

# VISUALIZACIÓN DE INCERTIDUMBRE

# Visualización de Incertidumbre

- La visualización de incertidumbre es uno de los aspectos más complicados en visualización de datos:
  - Cuando se observa un punto, relacionado a un dato, en una posición específica se tiende a interpretarlo como una representación precisa del verdadero valor del dato. Sin embargo, es posible que los datos registrados no representen precisamente los valores de los fenómenos; en otras palabras, el punto asociado a un dato puede estar localizado en otro lugar. Este fenómeno es común en ciencia y muestreo y por lo tanto en visualización de datos. De hecho, casi todos los conjuntos de datos disponibles poseen algún grado de incertidumbre y la decisión de cómo visualizarla (si es que se decide hacerlo) puede marcar una diferencia importante en como se percibe los datos y su significado.
- En ciencia, la utilización de barras de error o bandas de confidencia son dos formas comunes para representar la incertidumbre, pero su interpretación puede requerir de cierto nivel de conocimiento experto; a pesar de ello, estos métodos son precisos y aprovechan bien el espacio de visualización (p.ej., la utilización de barras de error permite mostrar en una sola visualización la incertidumbre de varios parámetros).
- Sin embargo, es deseable tener otras estrategias simples, fuertes e intuitivas de visualización de incertidumbre.

# Visualización de Incertidumbre

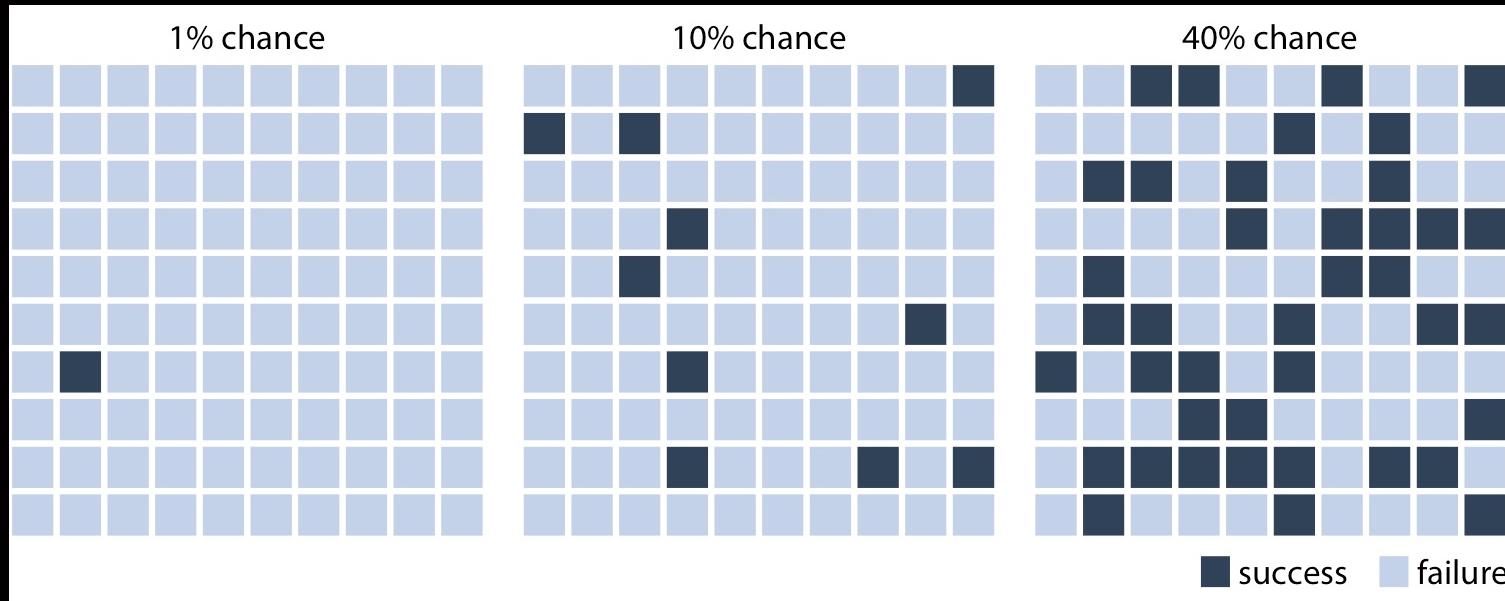
## ¿qué es incertidumbre?

- Algunos ejemplos de incertidumbre son:
  - No saber el resultado de un evento en el futuro (p.ej., el lanzamiento de una moneda).
  - No saber el momento exacto en que ocurrió un evento en el pasado (p.ej., asomarse a la ventana dos veces y ver un coche la primera vez y ya no verlo en la segunda).
- Desde el punto de vista matemático se utiliza la probabilidad (la posibilidad de que algo ocurra o sea). Aunque existe todo un cuerpo muy amplio de conocimiento acerca de probabilidad, en términos de visualización se pueden considerar las frecuencias relativas: la probabilidad de éxito está dado, aproximadamente, por la fracción de veces que un evento esperado se observa si se realizan varios ensayos aleatorios (p.ej., si un evento ocurre con una probabilidad de 10%, se puede pensar que ese evento se observa 1 de cada 10 ensayos).

# Visualización de Incertidumbre

## ¿cómo visualizar probabilidades?

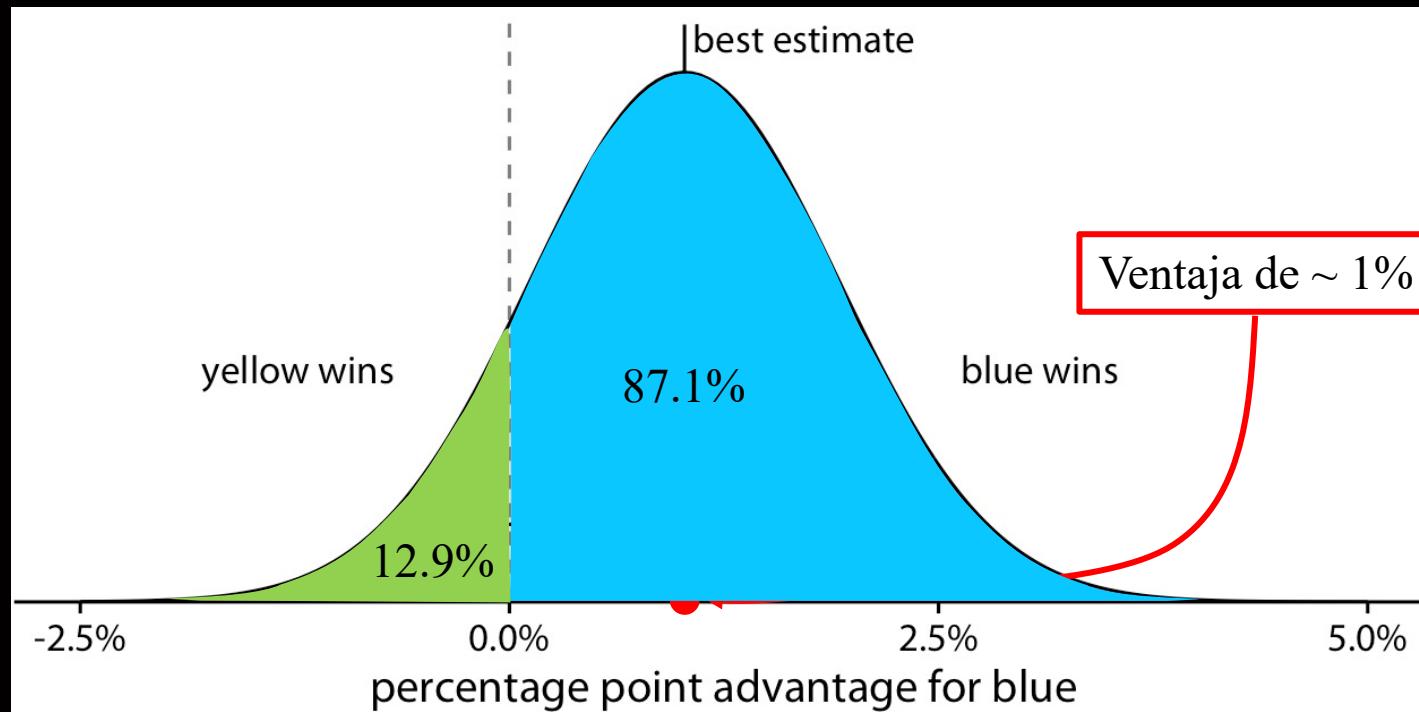
- Con base en las técnicas que hemos visto en el curso, se podría elegir visualizar los valores de probabilidad (un número) usando algún método de puntos o de barras. Sin embargo, estos métodos no dejan percibir la idea de probabilidad (frecuencia de eventos exitosos), de hecho, en general los humanos no somos muy buenos percibiendo el concepto de probabilidad.
- Alternativamente, se puede usar una técnica de visualización de resultados discretos:



# Visualización de Incertidumbre

## ¿cómo visualizar probabilidades?

- La visualización anterior funciona muy bien cuando los eventos son discretos, pero no funcionan cuando se trata de variables aleatorias con un número infinito de posibilidades.
- En ese caso, una alternativa es la utilización de distribuciones para mostrar posibles resultados:



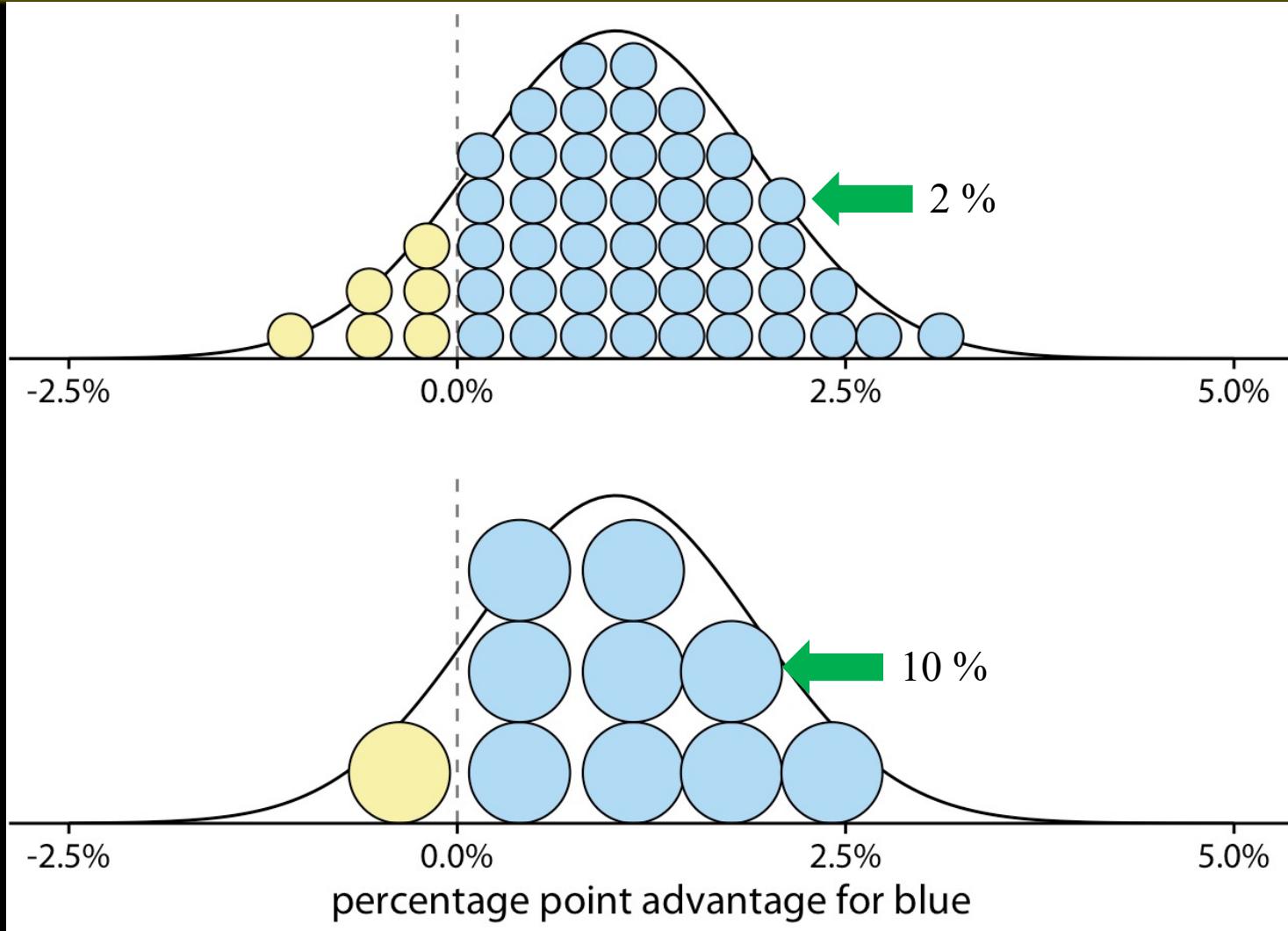
# Visualización de Incertidumbre

## ¿cómo visualizar probabilidades?

- Si se comparan las dos figuras anteriores, la primera figura da una mejor impresión de la incertidumbre, aunque las regiones rellenas de la segunda figura ofrece una mejor idea de las probabilidades de triunfo para ambos equipos.
- La investigación en percepción humana ha demostrado que somos mejores para percibir, contar y juzgar las frecuencias relativas de objetos discretos, siempre y cuando no se encuentren en grandes cantidades, que lo que somos juzgando los tamaños relativos de diferentes áreas.
- Se pueden combinar los tipos de resultados discretos de la primera figura con la distribución continua de la segunda si se dibuja una gráfica de puntos cuantiles. En este tipo de gráficas se subdivide el área total bajo la curva en unidades del mismo tamaño que se dibujan como círculos. Éstos se apilan de tal forma que su arreglo representa aproximadamente la distribución original de la curva.

# Visualización de Incertidumbre

## ¿cómo visualizar probabilidades?



# Visualización de Incertidumbre

## ¿cómo visualizar probabilidades?

- Las gráficas de puntos cuantiles deben usar un número pequeño o moderado de puntos; si hay muchos puntos, la gráfica se comienza a percibir como continua en lugar de hecha de puntos.
- En casos como la segunda gráfica mostrada antes, se hace un compromiso con la precisión matemática por una mejor percepción humana para obtener una mejor visualización (en particular, cuando no se tiene una audiencia especializada).
- Una visualización que es correcta matemáticamente pero que no se percibe apropiadamente no es útil en la práctica.

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- En estadística, el conjunto total de elementos (p.ej., ciudadanos) forman una población y un conjunto de ellos es una muestra de esa población. La población representa el estado verdadero del mundo y la muestra es una ventana hacia ese mundo.
- En general, nos interesan cantidades específicas que resumen propiedades importantes de la población. Las cantidades que describen a una población se les conoce como *parámetros* y son normalmente desconocidos.
- Para conocer esos parámetros se puede utilizar la muestra para realizar una conjetura sobre los valores verdaderos de esos parámetros (*estimaciones*). Por ejemplo, la media de una muestra es una estimación de la media poblacional (que es un parámetro). Las estimaciones de los valores de parámetros individuales se conocen como estimaciones puntuales (porque se pueden representar como un punto sobre la línea).

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- El número de observaciones individuales en una muestra se conoce como *tamaño de muestra*. De la muestra se pueden calcular algunos parámetros tales como su media y desviación los cuales, en general, difieren de los correspondientes parámetros de la población, en este caso la media y la desviación estándar. Si se repite el proceso de muestreo varias veces y se calculan los parámetros de la muestra, entonces la media resultantes de dichas muestras se distribuirán de acuerdo a la distribución del muestreo de la media. El ancho de la distribución de muestreo se conoce como *error estándar* y da información acerca de que tan precisos son las estimaciones (el error estándar provee una medida de incertidumbre asociada con la estimación del parámetro, en este caso la media). Como regla general, entre mayor sea el tamaño de la muestra, menor será el error estándar y por lo tanto la incertidumbre de la estimación.
- Es muy importante que no se confundan la desviación estándar con el error estándar: La desviación estándar es una propiedad de la población y dice que tanto esparcimiento hay entre las observaciones individuales que podemos realizar. Por otra parte, el error estándar informa acerca de que tan precisamente se ha determinado una estimación de parámetro.

# Visualización de Incertidumbre

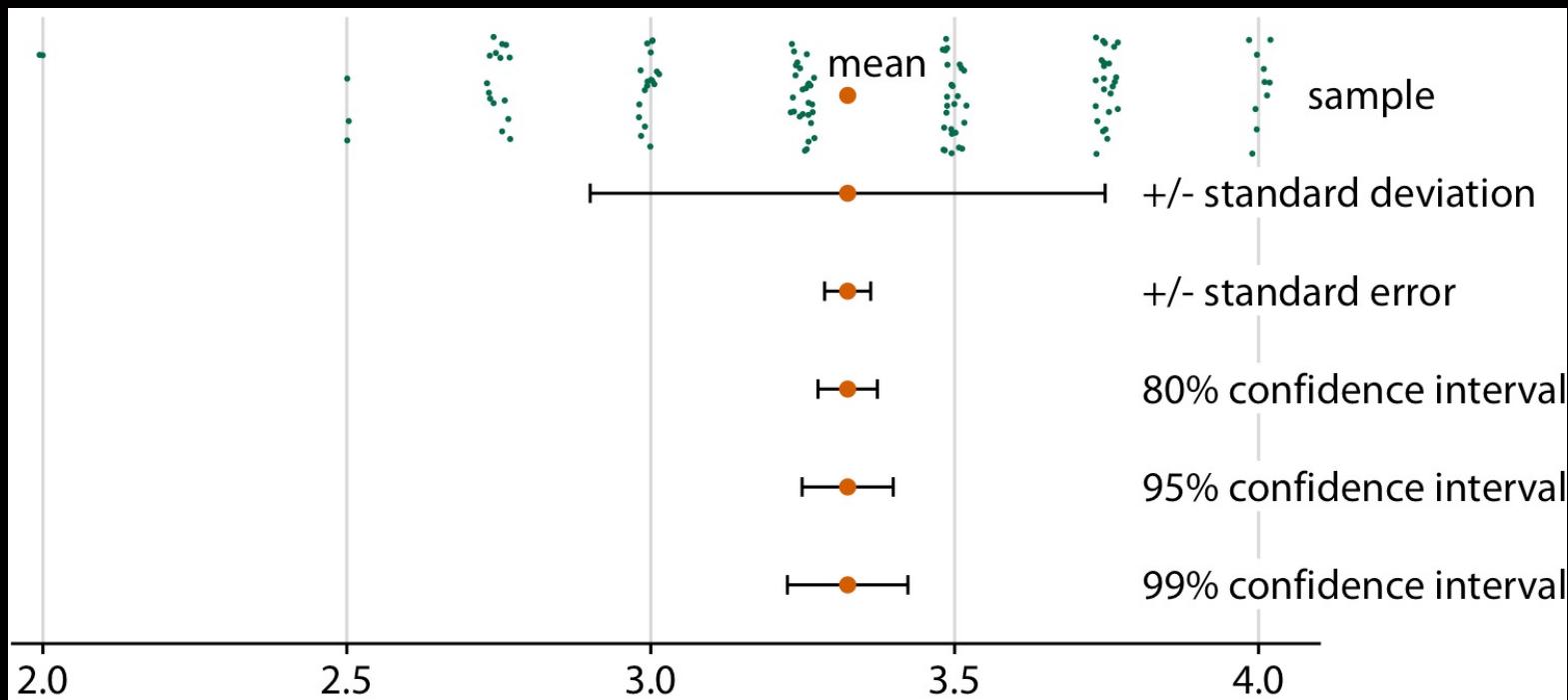
## Incertidumbre de Estimaciones Puntuales

- Todos estadísticos usan muestras para calcular estimaciones de parámetros y sus incertidumbres. Sin embargo, los estadísticos se encuentran divididos en cuanto a como se realizan estos cálculos y se pueden considerar dos escuelas: Bayesianos y Frecuencistas. Los Bayesianos asumen que hay algún conocimiento previo acerca del mundo y usan la muestra para actualizar ese conocimiento. Por otra parte, los Frecuentistas intentan hacer afirmaciones precisas acerca del mundo sin tener un conocimiento previo a la mano.
- Afortunadamente, ambas escuelas típicamente usan las mismas estrategias.

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- Los Frecuentistas comúnmente visualizan la incertidumbre con barras de error. Aún cuando las barras de error son útiles para mostrar incertidumbre, este tipo de visualización pueden llevar a confundir el error estándar, la desviación estándar u otras medidas de variación.



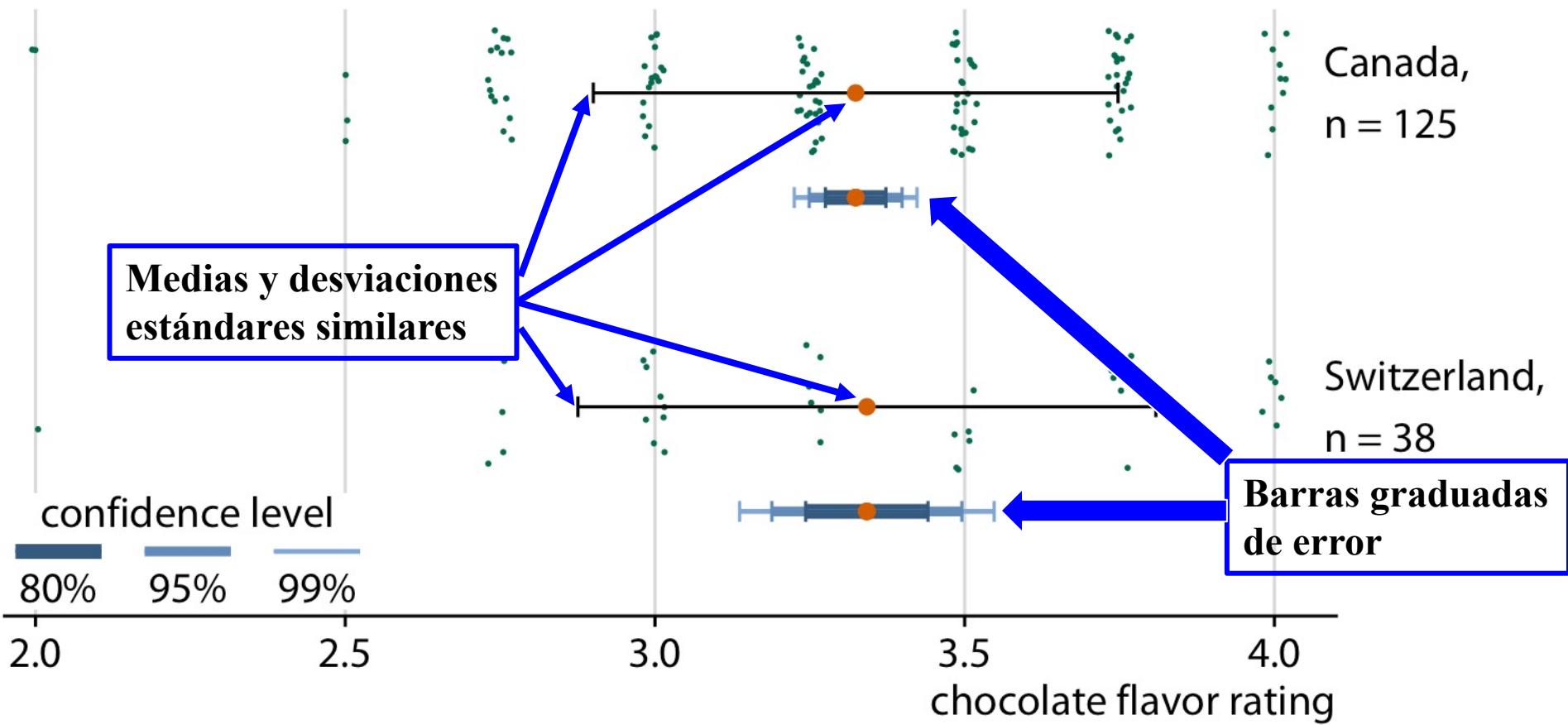
# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- El error estándar es aproximadamente igual a la relación entre la desviación estándar de la muestra con la raíz cuadrada del tamaño de la muestra. Los intervalos de confianza se calculan multiplicando el error estándar con valores constantes pequeños. Por ejemplo, un intervalo de confidencia de 95% se extiende aproximadamente a dos veces el error estándar en ambas direcciones de la media. Por lo tanto, muestras más grandes tienden a tener errores estándar e intervalos de confianza más estrechos, aún cuando su desviación estándar es la misma.

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

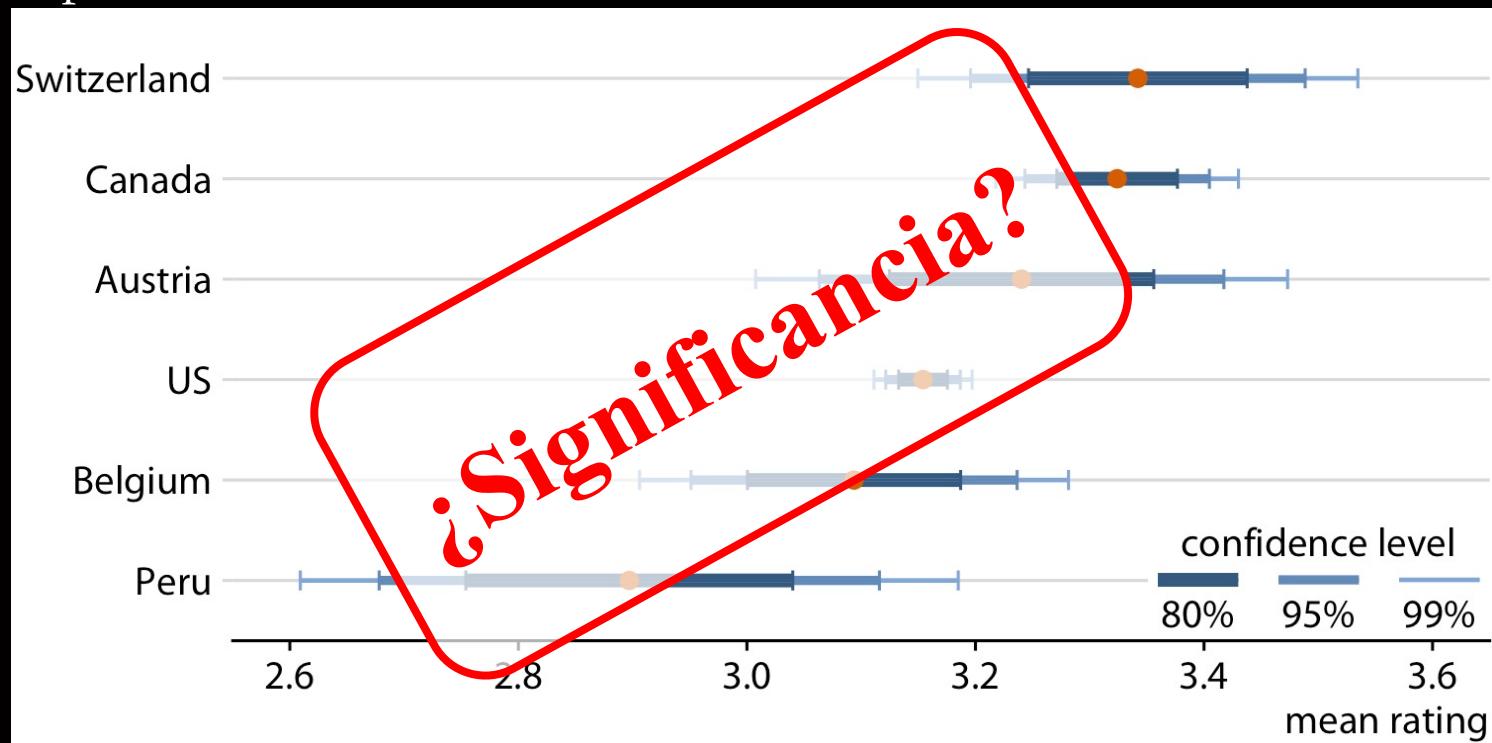


# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- Las barras de error son convenientes porque permiten mostrar todas las estimaciones con sus incertidumbres al mismo tiempo. Por lo tanto, son populares en publicaciones científicas donde el objetivo principal es, usualmente, conllevar una gran cantidad de información a una audiencia experta.

Medias y sus intervalos de confianza asociados de la clasificación de sabores de barras de chocolate de seis países productores.



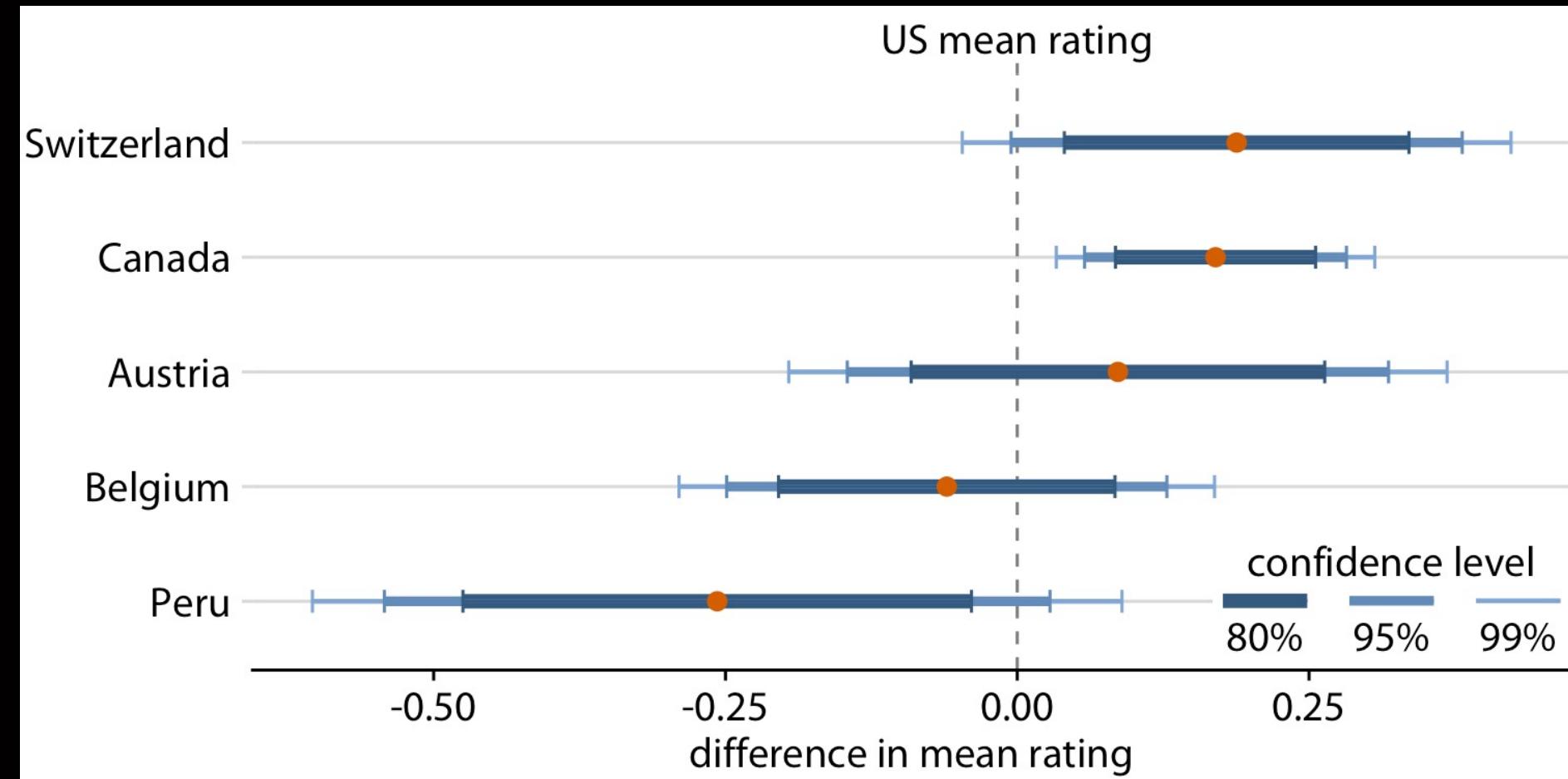
# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- La palabra significancia es un término estadístico que se refiere al nivel de confidencia para rechazar la suposición de que la diferencia observada fue ocasionada por un muestreo aleatorio.
- Aunque varios libros de texto y tutoriales de estadística recomiendan formas de juzgar significancia a partir de que tanto se traslapan o no las barras de error, estas recomendaciones no son confiables y es mejor ignorarlas.
- La forma correcta de juzgar si hay diferencias en las medias es calcular los intervalos de confianza para las diferencias. Si esos intervalos de confianza excluyen el cero, entonces la diferencia es significativa a un cierto nivel de confianza.

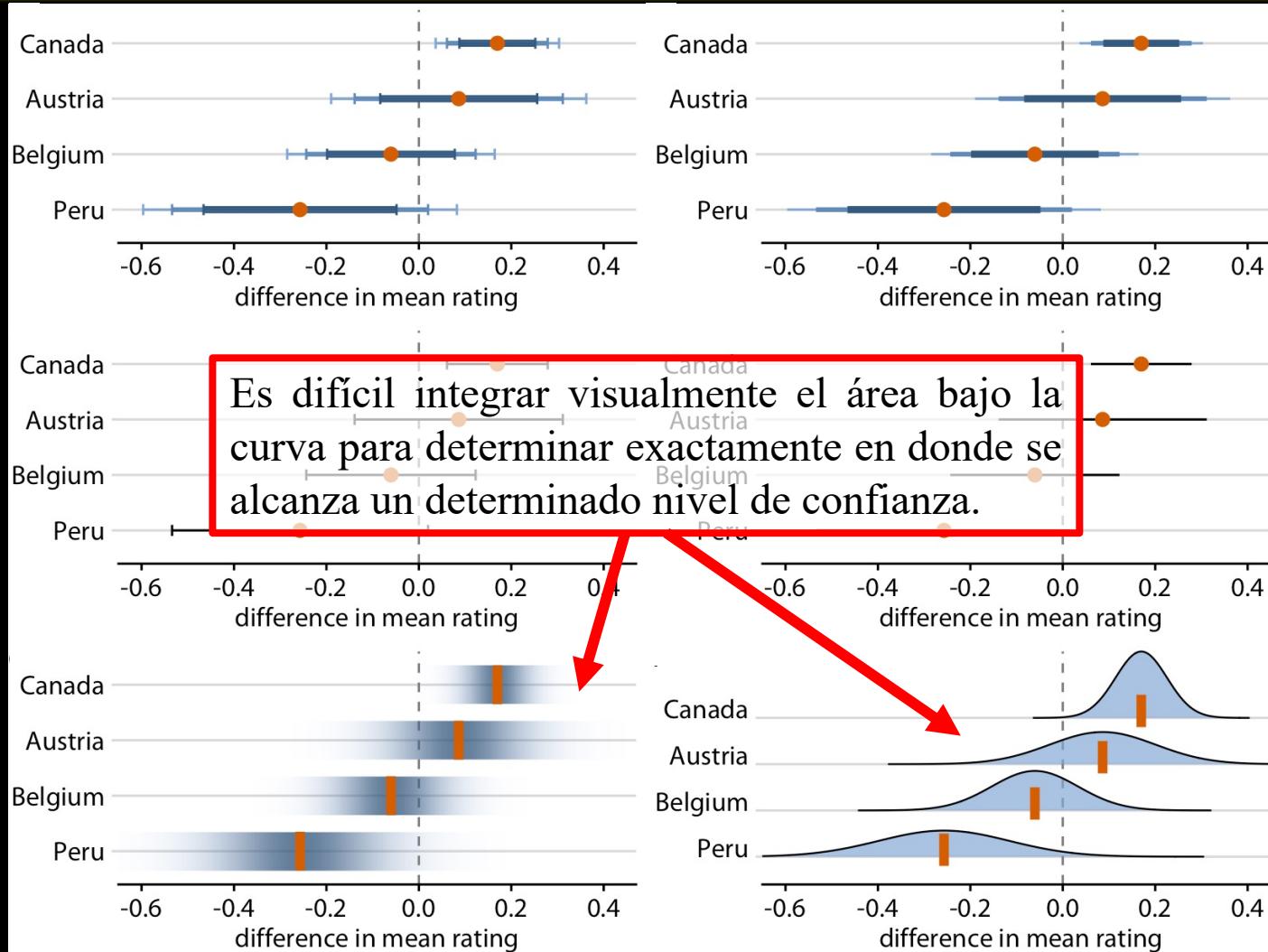
# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales



# Visualización de Incertidumbre

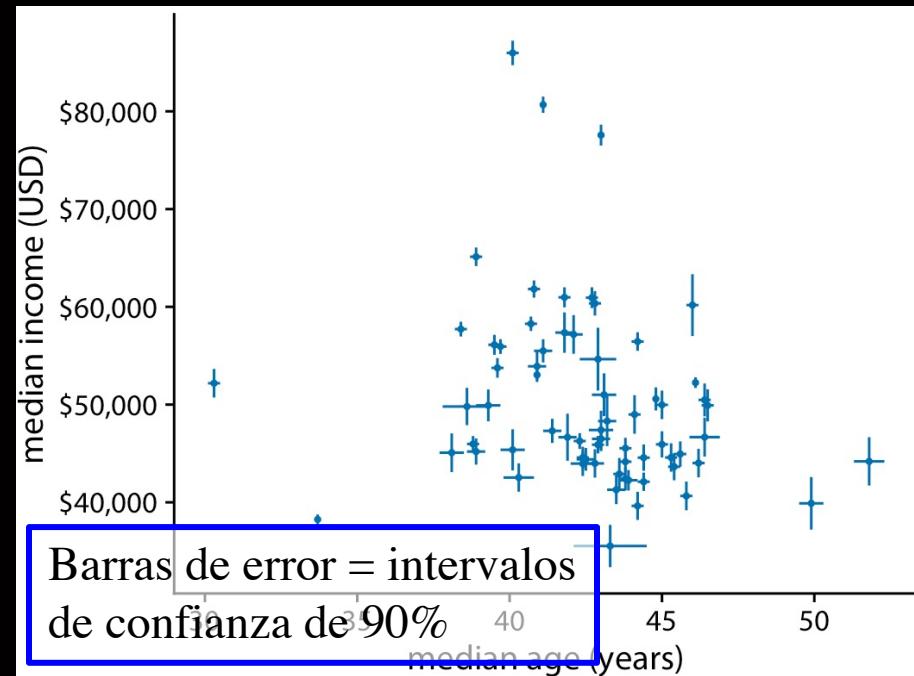
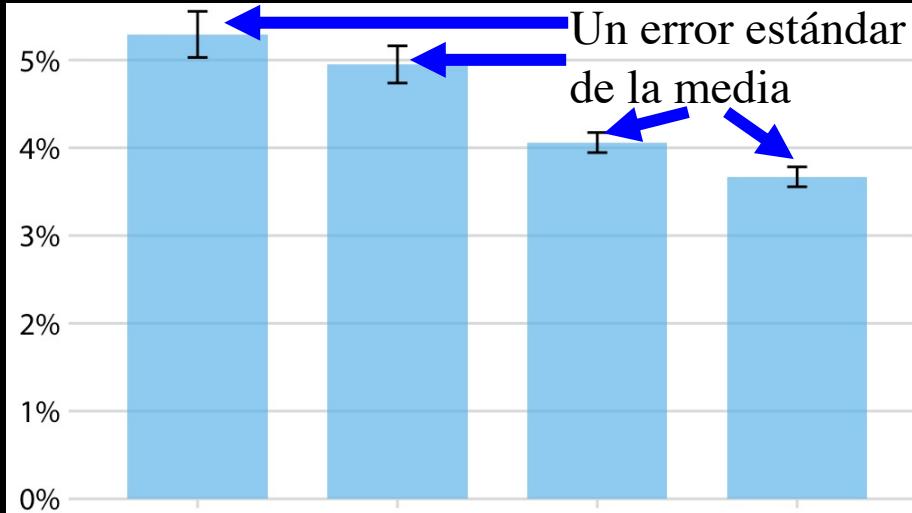
## Diferentes Formas de Visualizar Incertidumbre



# Visualización de Incertidumbre

## Diferentes Formas de Visualizar Incertidumbre

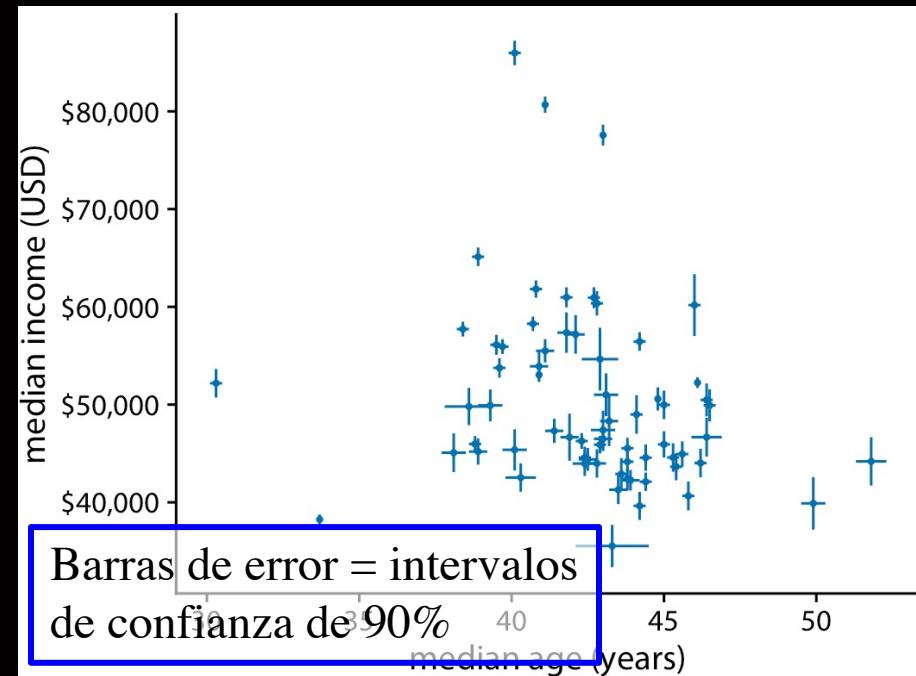
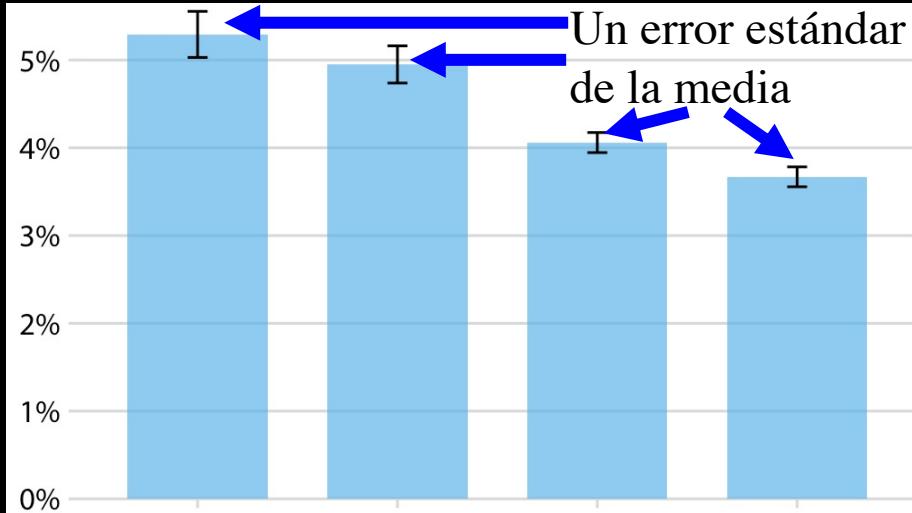
- Para figuras 2D simples, las barras de error poseen una ventaja importante sobre visualizaciones de incertidumbre más complejas: se pueden combinar con otras formas de gráficas. Para casi cualquier forma de visualización se puede agregar una indicación de incertidumbre usando barras de error.



# Visualización de Incertidumbre

## Diferentes Formas de Visualizar Incertidumbre

- Para figuras 2D simples, las barras de error poseen una ventaja importante sobre visualizaciones de incertidumbre más complejas: se pueden combinar con otras formas de gráficas. Para casi cualquier forma de visualización se puede agregar una indicación de incertidumbre usando barras de error.



# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- Frecuentistas evalúan incertidumbre con intervalos de confianza y Bayesianos calculan distribuciones posteriores e intervalos de credibilidad.
- La distribución posterior Bayesiana dice que tan probable son las estimaciones de parámetros específicos dados los datos de entrada. El intervalo de credibilidad indica un rango de valores dentro del cual el valor del parámetro se espera con cierta probabilidad, como resultado de la distribución posterior. Por ejemplo, un intervalo de credibilidad de 95% corresponde al 95% central de la distribución posterior. El valor verdadero del parámetro tiene un 95% de caer en el 95% del intervalo de credibilidad.
- Un intervalo de credibilidad Bayesiano indica la probabilidad en donde se encuentra el parámetro verdadero, mientras que el intervalo de confidencia Frecuentista indica donde es probable que el parámetro verdadero no se encuentre.

# Visualización de Incertidumbre

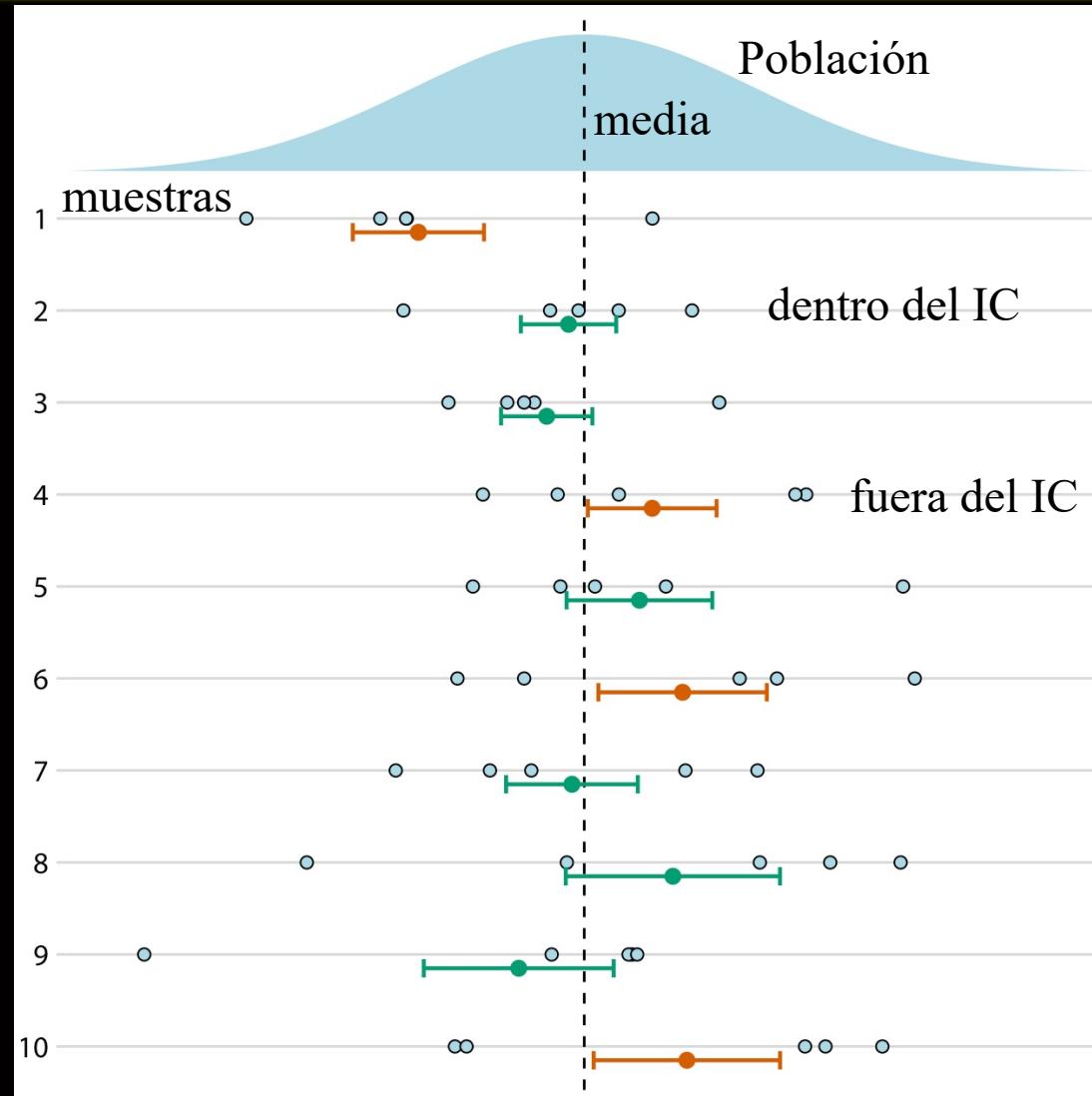
## Incertidumbre de Estimaciones Puntuales

- Bajo la perspectiva Bayesiana, se pueden usar los datos y el conocimiento previo acerca del sistema bajo estudio (el prior) para calcular una distribución de probabilidad (el posterior) que indica donde se espera que el valor verdadero del parámetro se encuentre.
- Bajo la perspectiva Frecuentista, primero se hace una suposición que se pretende refutar. Esta suposición se conoce como la *hipótesis nula* y típicamente es simplemente la suposición de que el parámetro es igual a cero (no existe diferencia entre dos condiciones). Después se calcula la probabilidad de que un muestreo aleatorio genere datos similares a lo que fue observado si la hipótesis nula hubiese sido verdadera. El intervalo de confianza es una representación de esta probabilidad. Si un intervalo de confianza excluye el valor del parámetro bajo la hipótesis nula (el valor cero), entonces se puede rechazar la hipótesis nula a ese nivel de confianza. Alternativamente, se puede considerar que el intervalo de confianza es un intervalo que captura el valor verdadero del parámetro con la probabilidad especificada bajo un muestreo repetido. Por lo tanto, si el valor verdadero del parámetro fuese cero, un intervalo de confianza de 95% solo excluiría el valor cero en 5% de las muestras analizadas.

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

Los intervalos de confianza muestran una desviación estándar ( $\sim 68\%$ ). Si se realizan varios muestreos, la media verdadera se incluirá en los intervalos de confianza 68% de las veces.



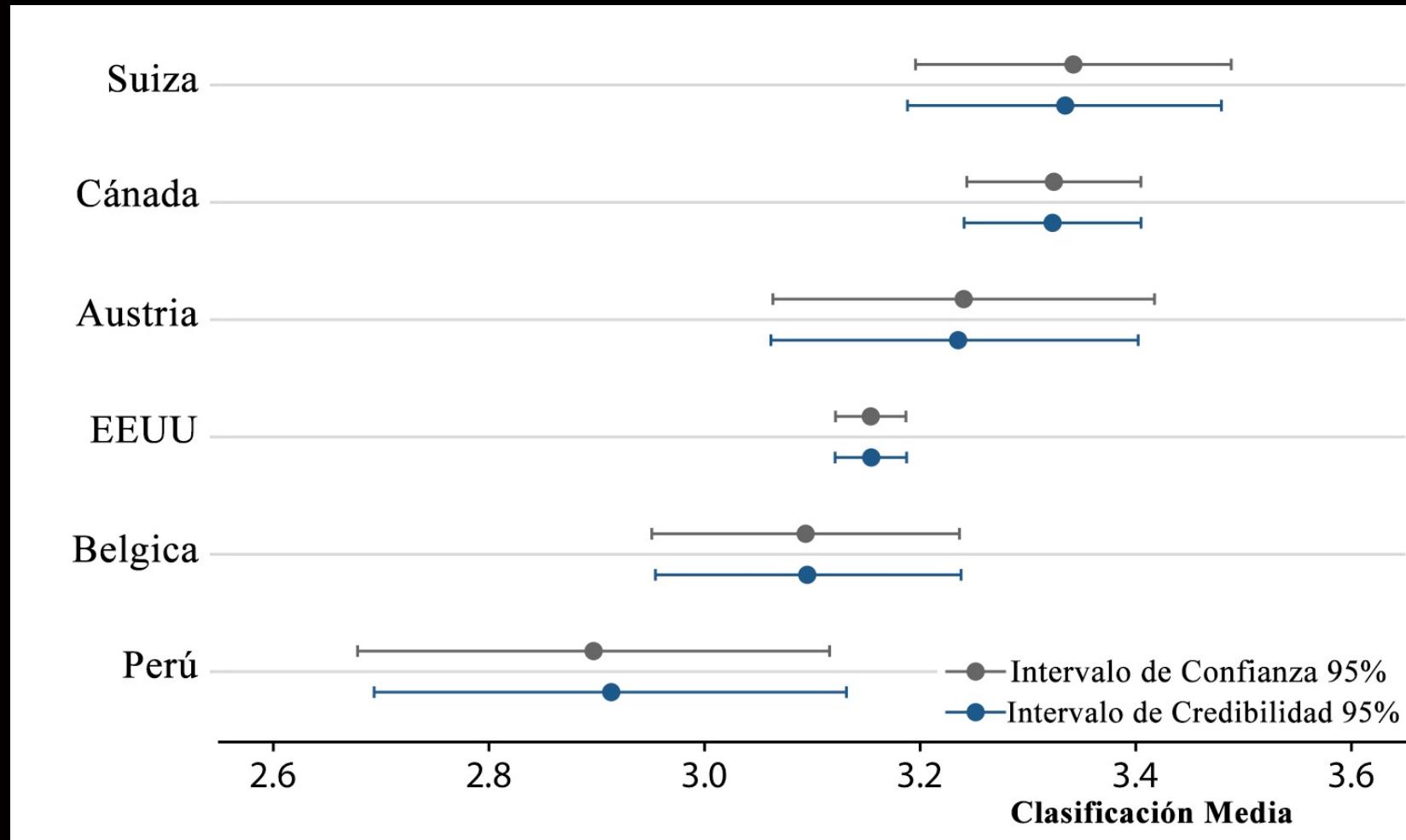
# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- Un intervalo de credibilidad Bayesiano hace una afirmación acerca del valor verdadero del parámetro y un intervalo de confianza Frecuentista hace una afirmación acerca de la hipótesis nula.
- En la práctica, las estimaciones Bayesianas y Frecuentistas son bastante similares. Una ventaja conceptual del pensamiento Bayesiano es que enfatiza considerar la magnitud de un efecto, los Frecuentistas, por otro lado, enfatiza una perspectiva binaria de un efecto ya sea que exista o no.
- Un intervalo de credibilidad Bayesiano contesta la pregunta “¿En donde se espera que se encuentre el valor verdadero del parámetro?”.
- Un intervalo de confianza Frecuentista contesta la pregunta “¿con cuanta confidencia estamos seguros de que el valor verdadero del parámetro no es cero?”.

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales



# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

- Un intervalo de credibilidad Bayesiano hace una afirmación acerca del valor verdadero del parámetro y un intervalo de confianza Frecuentista hace una afirmación acerca de la hipótesis nula.
- Un intervalo de credibilidad Bayesiano contesta la pregunta “¿En donde se espera que se encuentre el valor verdadero del parámetro?”.
- Un intervalo de confianza Frecuentista contesta la pregunta “¿con cuanta confidencia estamos seguros de que el valor verdadero del parámetro no es cero?”.
- En la práctica, las estimaciones Bayesianas y Frecuentistas son bastante similares. Una ventaja conceptual del pensamiento Bayesiano es que enfatiza considerar la magnitud de un efecto, los Frecuentistas, por otro lado, enfatiza una perspectiva binaria de un efecto ya sea que existe o no.

# Visualización de Incertidumbre

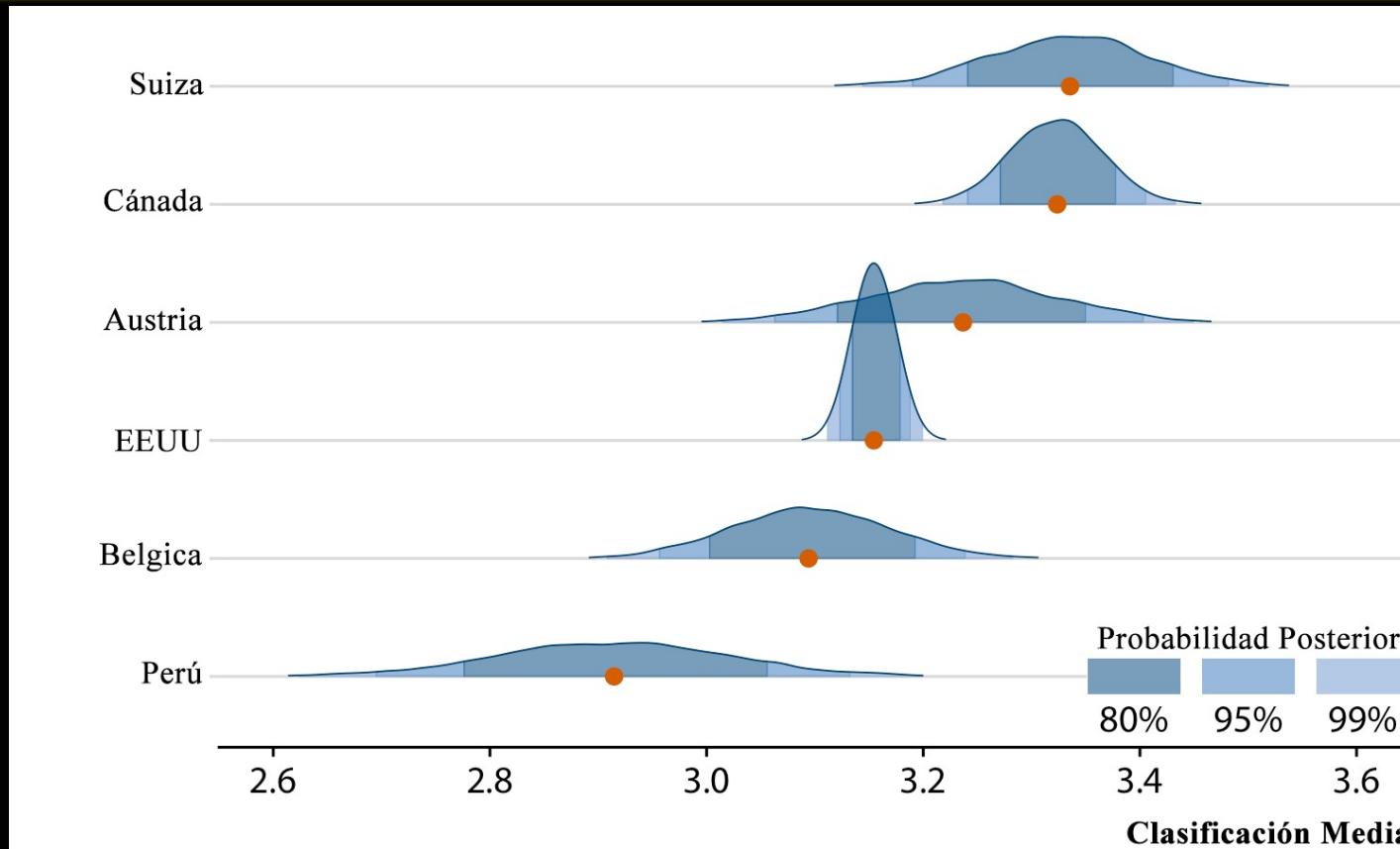
## Incertidumbre de Estimaciones Puntuales

---

- El objetivo principal de la estimación Bayesiana es la de obtener la distribución posterior y por ello es común que los Bayesianos visualicen la distribución entera en lugar de simplificarla en un intervalo creíble.
- Así que para la visualización de datos, todos los métodos para visualizar distribuciones son válidos para visualizar distribuciones posteriores, tales como histogramas, gráficas de densidad, violines y gráficas de crestas.

# Visualización de Incertidumbre

## Incertidumbre de Estimaciones Puntuales

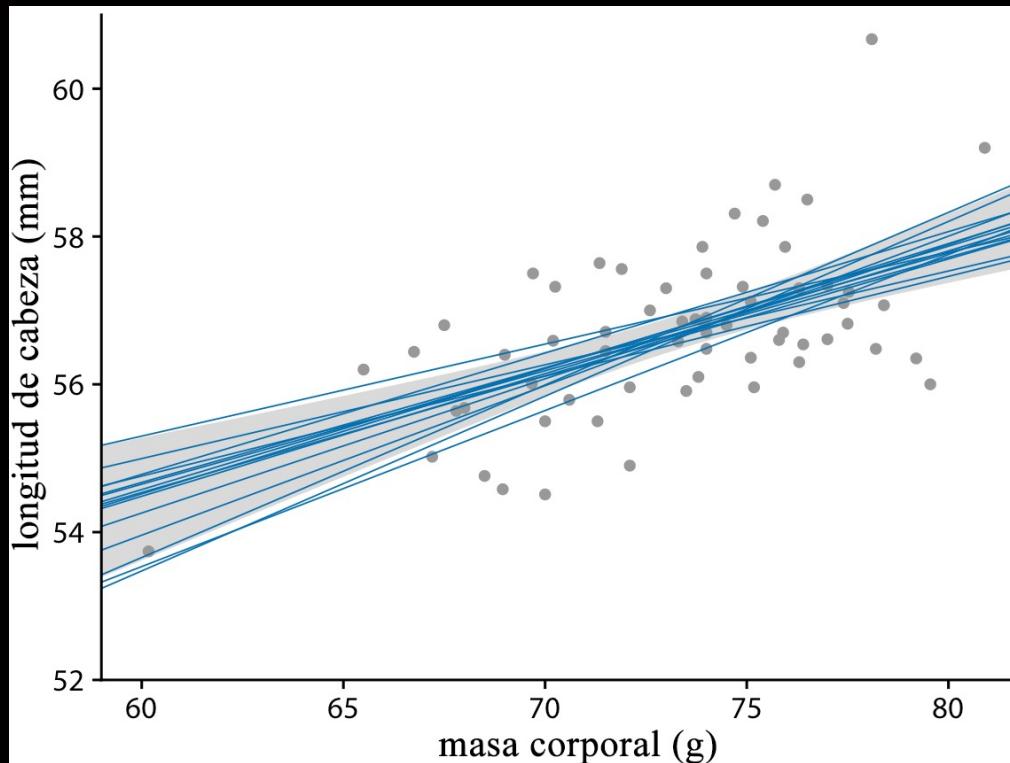


Gráfica de Crestas con sombreado bajo la curva para indicar regiones definidas de probabilidades posteriores. Se pueden usar barras de error graduadas debajo de las distribuciones en lugar de sombreado (*half eyes*). Gráficas de violín con barras de error (*Gráficas de ojo*).

# Visualización de Incertidumbre

## Incertidumbre de Ajustes de Curva

- Cuando vimos métodos para visualizar tendencias nos preocupamos por mostrar únicamente la tendencia principal. Estas tendencias también tienen incertidumbre que se acostumbra mostrar usando una *banda de confianza*. Una banda de confianza provee un rango de diferentes líneas de ajuste que podrían ser compatibles con los datos.



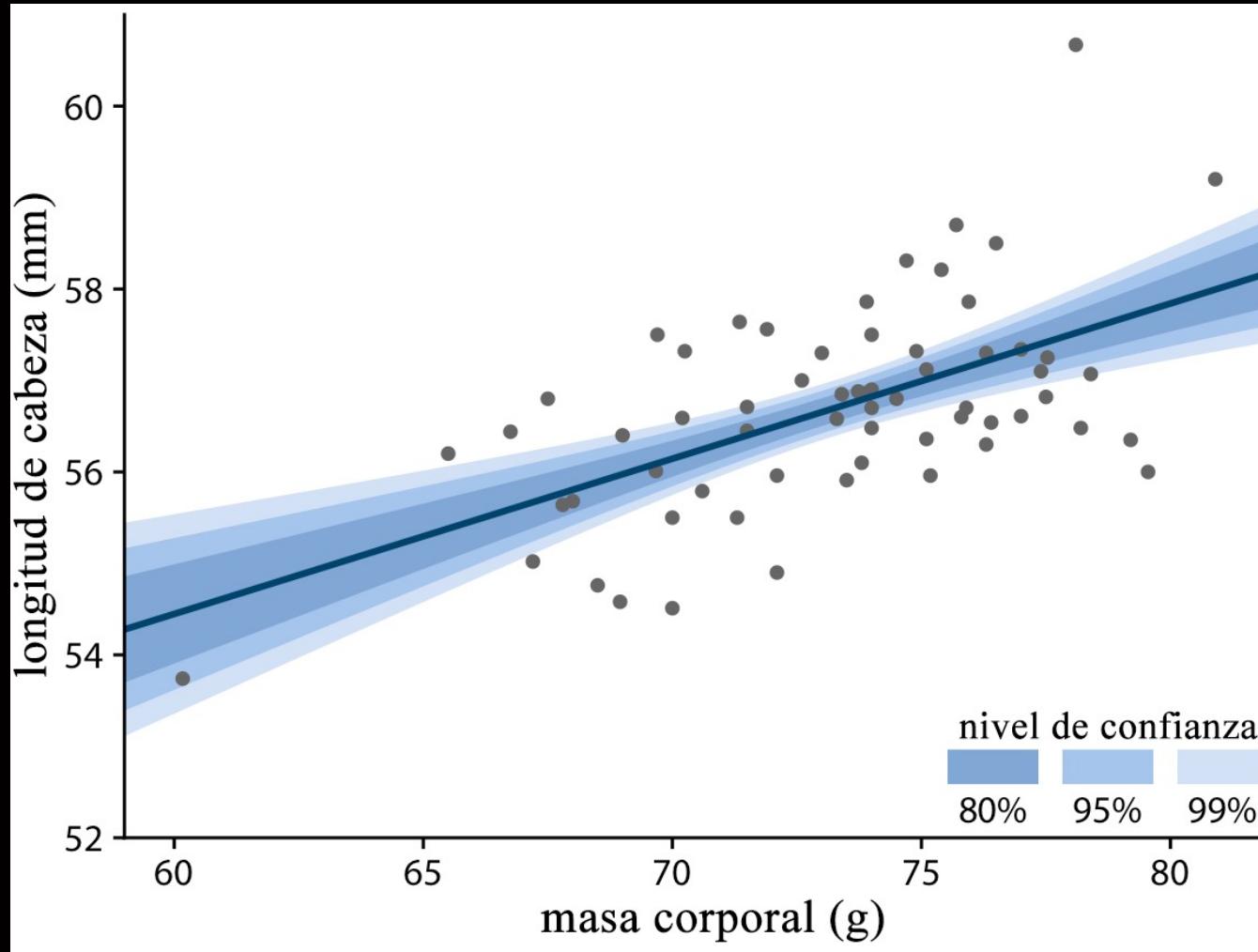
# Visualización de Incertidumbre

## Incertidumbre de Ajustes de Curva

- Para dibujar una banda de confianza se necesita especificar un nivel de confianza y puede ser útil resaltar diferentes niveles de confianza. Esto resulta en la banda de confianza graduada, la cual muestra varios niveles de confianza en la misma visualización. Una banda de confianza graduada incrementa la apreciación de incertidumbre y obliga confrontar la posibilidad de que los datos puedan soportar diferentes líneas alternativas de tendencia.

# Visualización de Incertidumbre

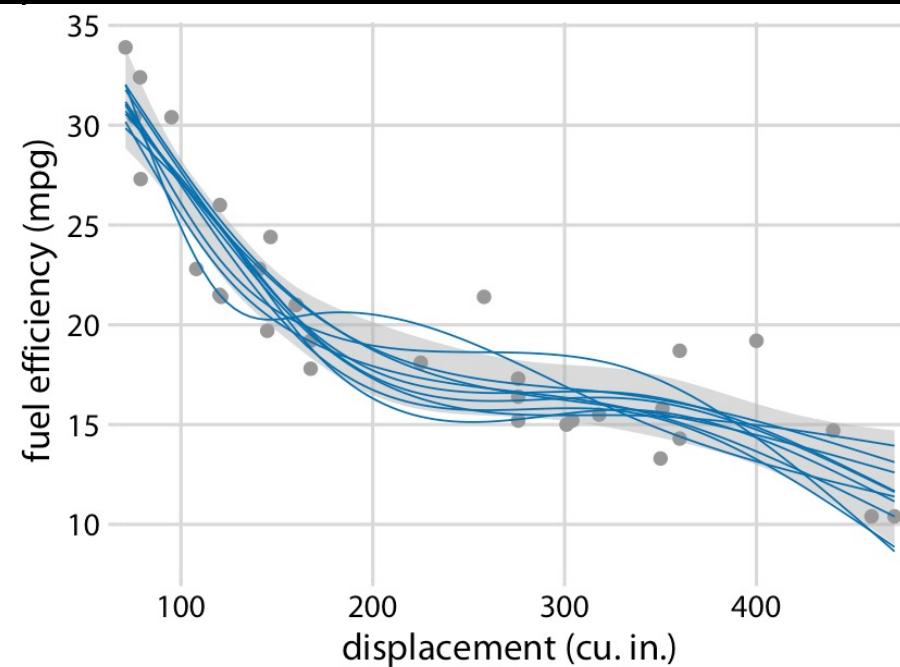
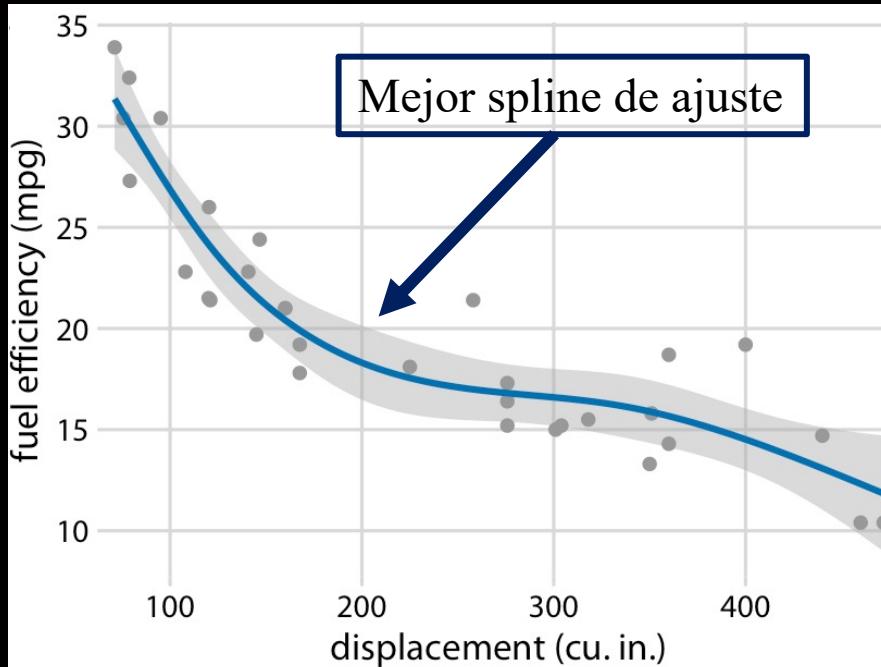
## Incertidumbre de Ajustes de Curva



# Visualización de Incertidumbre

## Incertidumbre de Ajustes de Curva

- También se pueden dibujar bandas de confianza para ajustes de curvas no-lineales. Dichas bandas de confianza se ven bien pero pueden ser difíciles de interpretar.
- Es importante hacer notar que las bandas de confianza se crean a partir de varias líneas de tendencia alternativas.



# Visualización de Incertidumbre

## Gráficas de Resultados Hipotéticos

- Todas las visualizaciones estáticas de incertidumbre sufren del problema de que los observadores pueden interpretar algún aspecto de la visualización de incertidumbre como una característica determinística de los datos (un error de construcción determinista).
- Este problema se puede evitar si se utiliza animación para ciclar a través de un número de diferentes gráficas que son igual de probables.
- Este tipo de visualización se conoce como gráfica de resultado hipotético (hypothetical outcome plot – HOP). Gráficas de este tipo pueden ser muy efectivas en un ambiente en línea donde se pueden utilizar medios como MP4 o GIF.