Excelencia gráfica

Visualización Científica
6 de septiembre de 2016
Eduardo Castro-Nallar, PhD
Center for Bioinformatics and Integrative Biology
www.cbib.cl
www.castrolab.org

Excelencia gráfica

- Conjunto de principios que nos ayudan a inducir al lector a pensar sobre la substancia de los datos presentados

¿Por qué no simplemente usamos tablas?

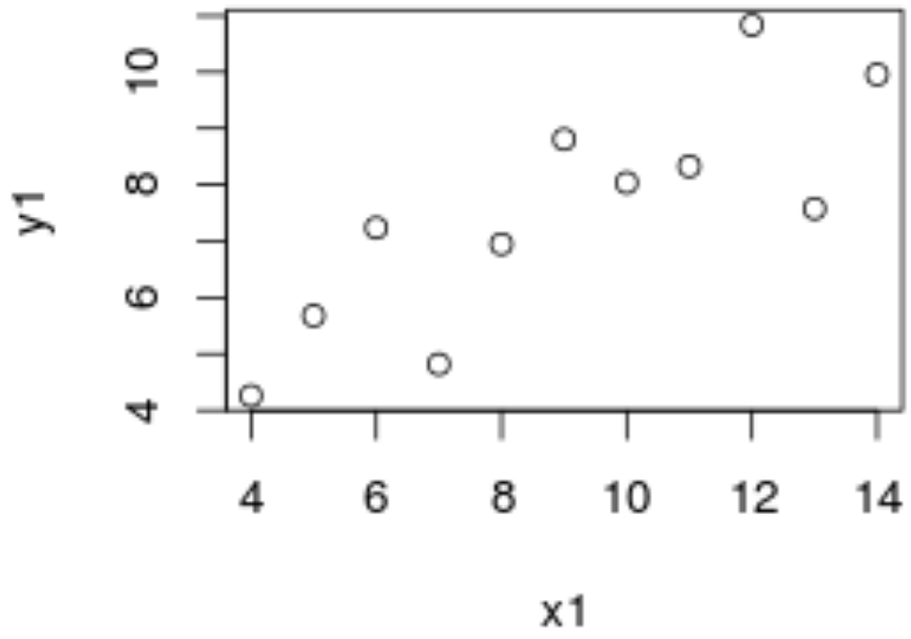
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

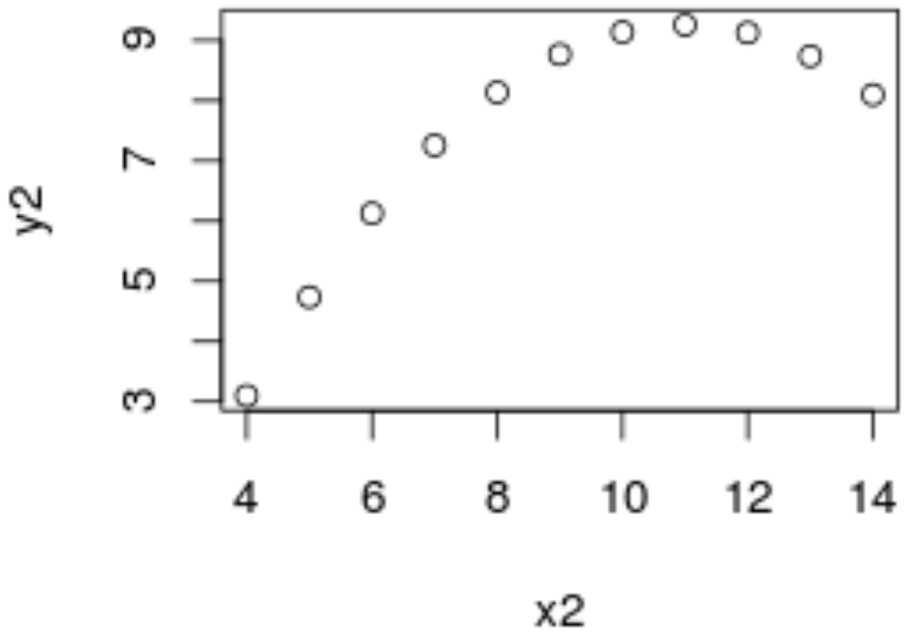
E. R. Tufte. The Visual Display of Quantitative Information, 2nd Edition. Graphics Press, Cheshire, Connecticut, 2001.

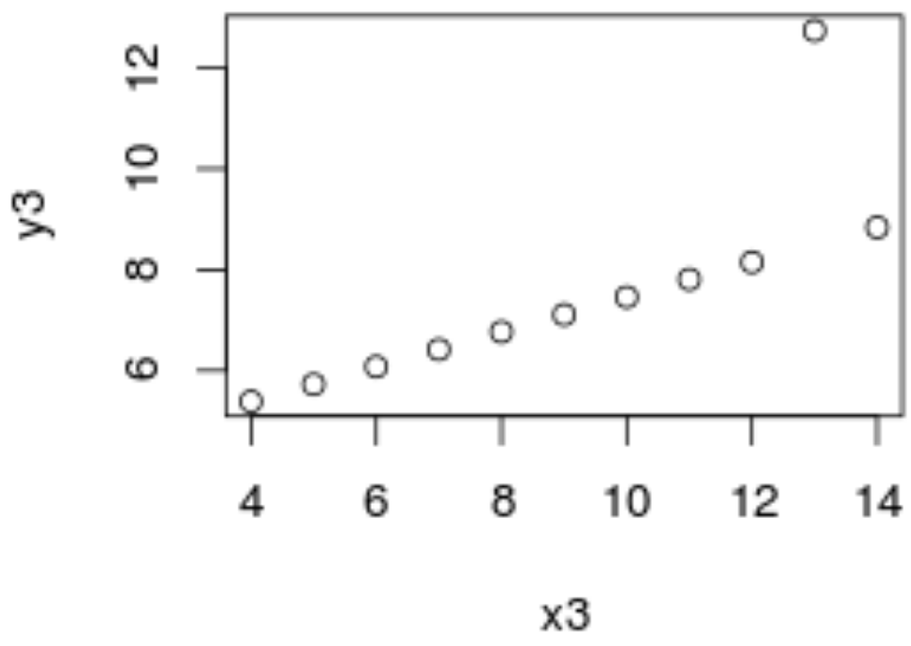
¿Por qué no simplemente usamos tablas?

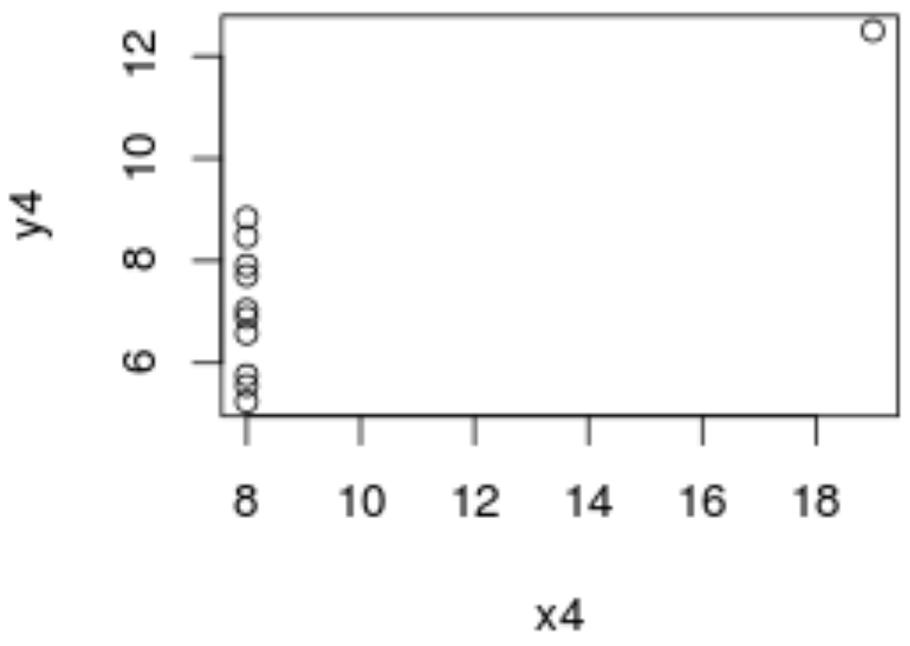
- El promedio de los valores x es 9 para cada dataset
- El promedio de los valores y es 7.50 para cada dataset
- La varianza de x es 11 y la de y es 4.12
- El coeficiente de correlación entre x e y es 0.816 para cada dataset
- La ecuación de la recta para cada dataset es y = 0.5x + 3

¿Podemos descubrir algo si visualizamos los datos?







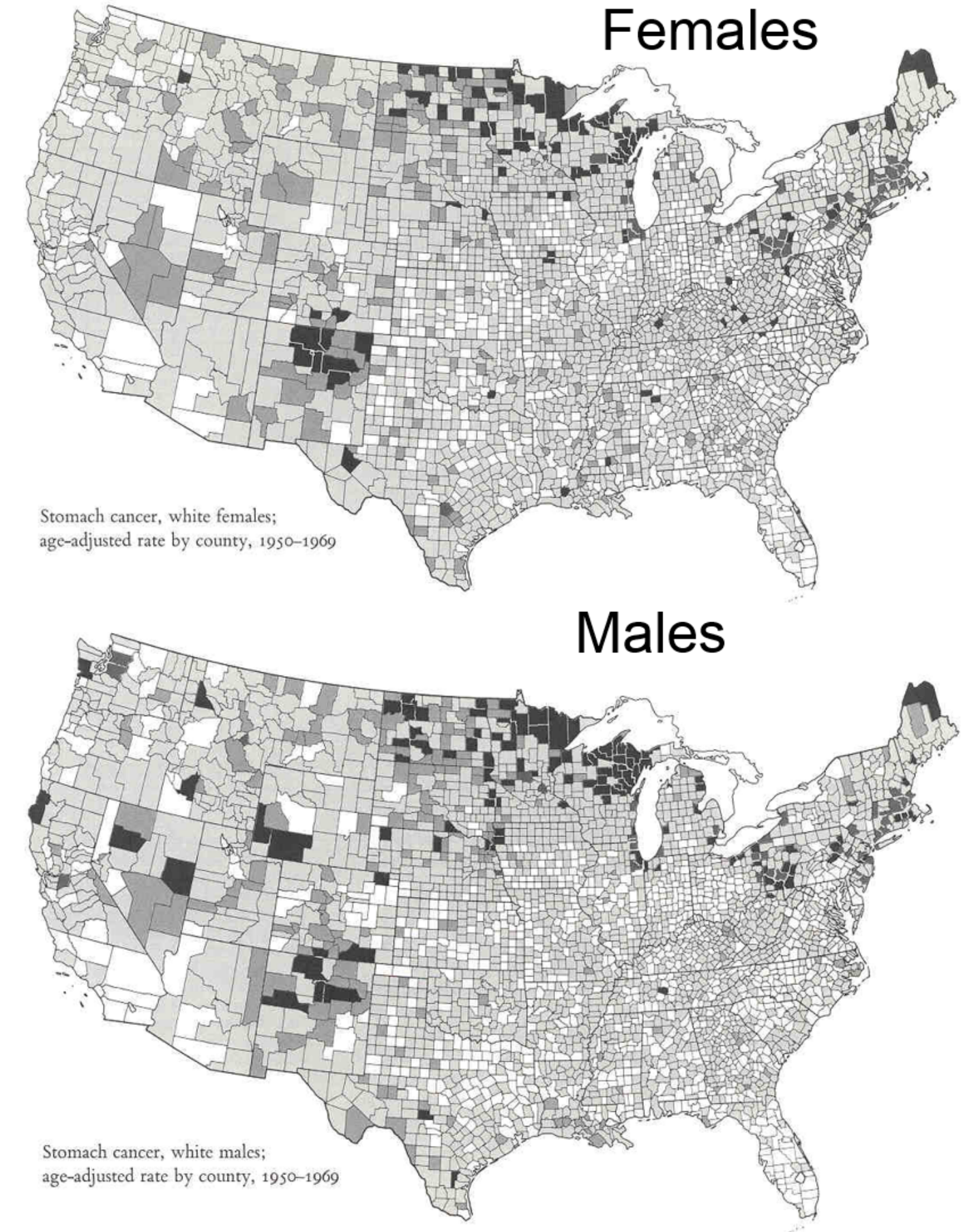


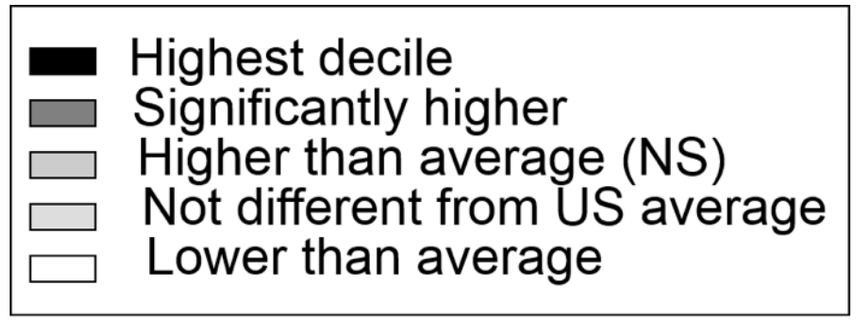
E. R. Tufte. The Visual Display of Quantitative Information, 2nd Edition. Graphics Press, Cheshire, Connecticut, 2001.

Excelencia gráfica

- Mostrar los datos
- Inducir al lector a considerar la sustancia de los resultados en vez de la metodología, el diseño, o cualquier otra cosa
- Evitar distorsionar lo que los datos pueden decir
- Presentar números eficientemente
- Mostrar gran cantidad de datos de manera coherente
- Estimular comparaciones
- Revelar los datos a distintos niveles de detalle y complejidad
- El gráfico tiene que tener un propósito claro: descripción, exploración, tabulación, etc.
- Integrar el gráfico con los métodos estadísticos usados para analizar los datos

E. R. Tufte. The Visual Display of Quantitative Information, 2nd Edition. Graphics Press, Cheshire, Connecticut, 2001.





Tasas de cáncer por condado en EE.UU

- Tasa de mortalidad para cáncer de estómago
- Cada mapa muestra ~21000 puntos
- Eficiente, coherente, efectivo

E. R. Tufte. The Visual Display of Quantitative Information, 2nd Edition. Graphics Press, Cheshire, Connecticut, 2001.

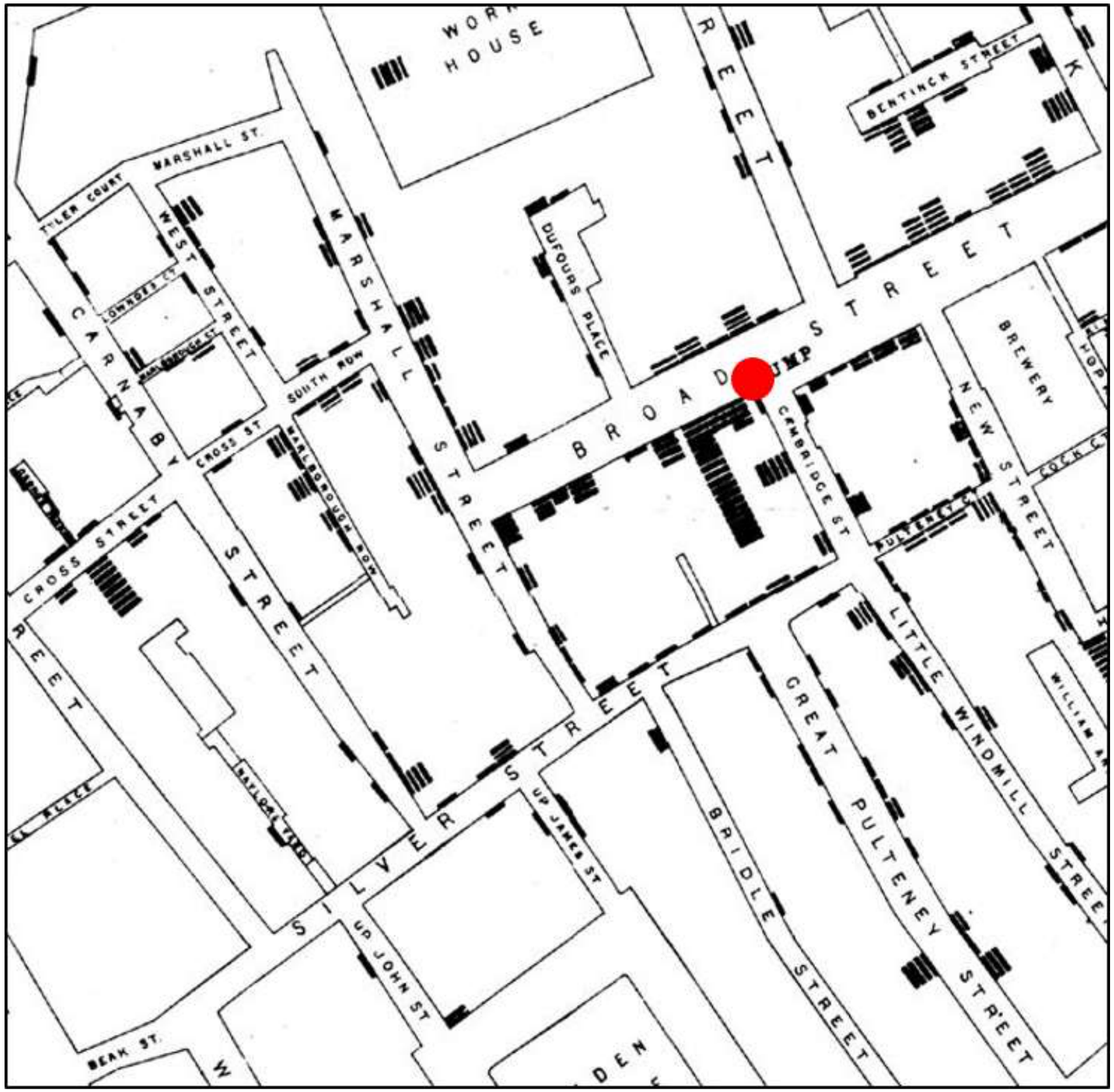


Figura por John Snow en 1854

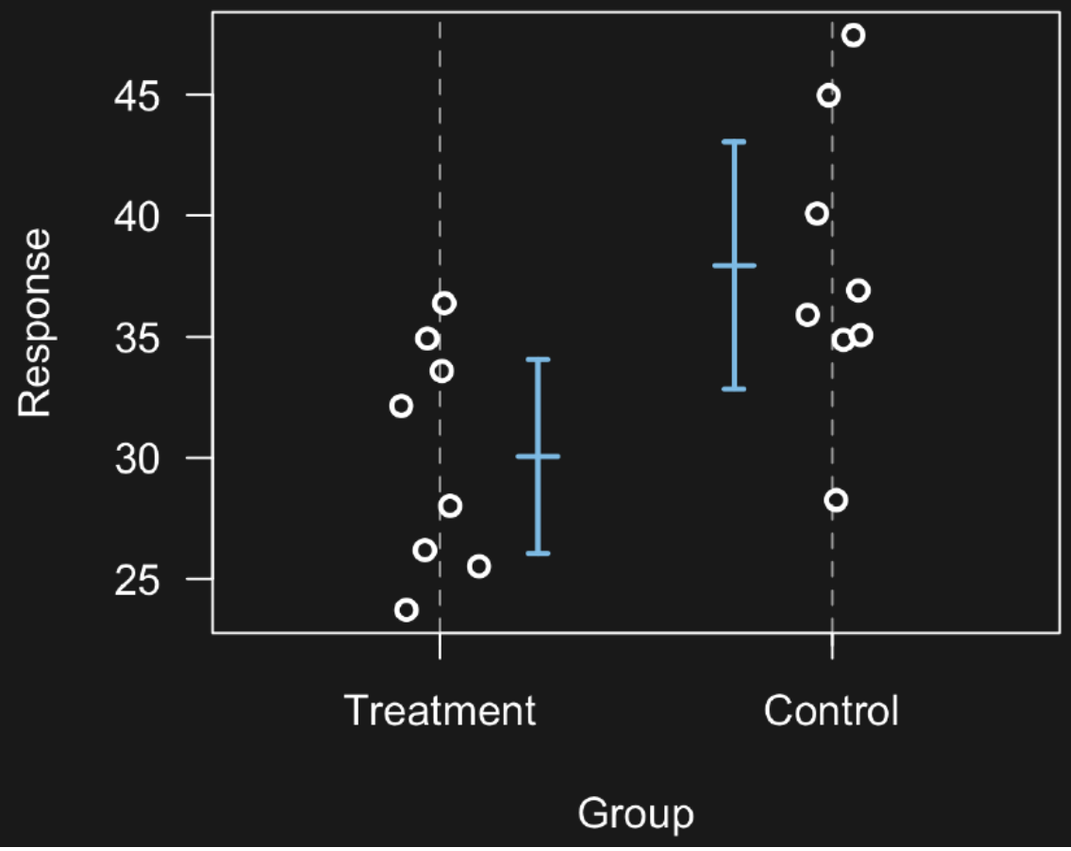
- Cada disco negro representa un caso de cólera
- El punto rojo representa un grifo de agua en la calle Broad

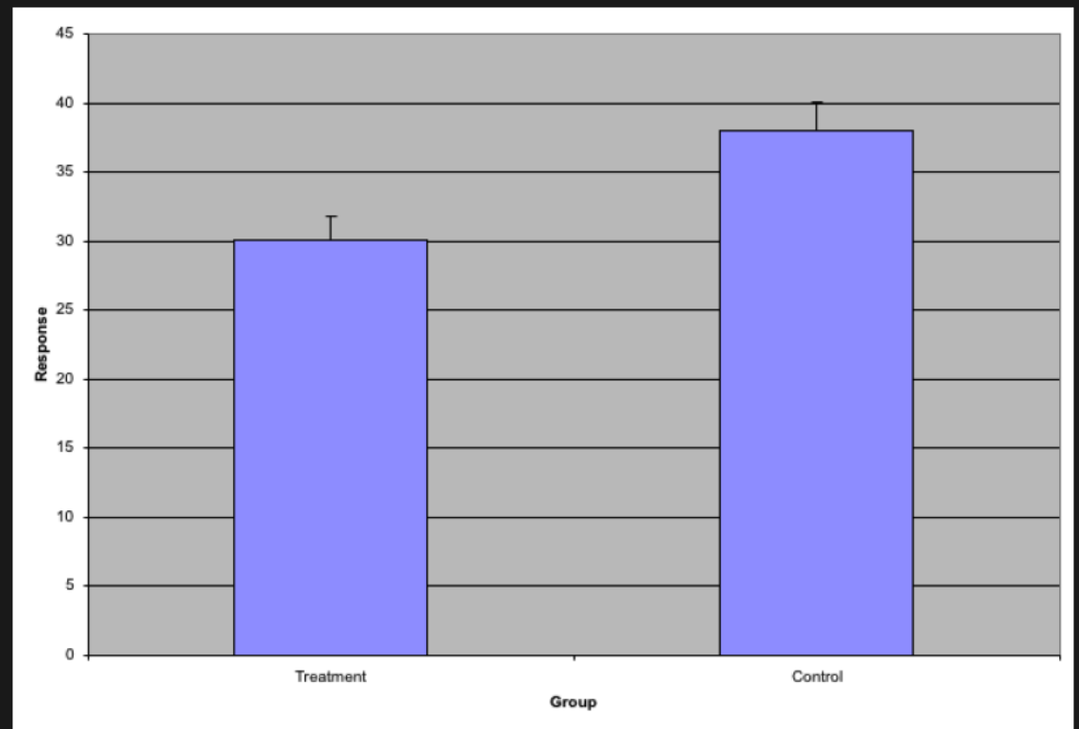
E. R. Tufte. The Visual Display of Quantitative Information, 2nd Edition. Graphics Press, Cheshire, Connecticut, 2001.

Principios a seguir

- Ser exacto y claro
- La mayor cantidad de información posible sin obscurecer el mensaje
- Ciencia y no ventas, i.e., nada de 3D
- Rigurosidad con las cifras significativas
- Siempre mostrar incertidumbre en mediciones

¿Qué NO hacer?

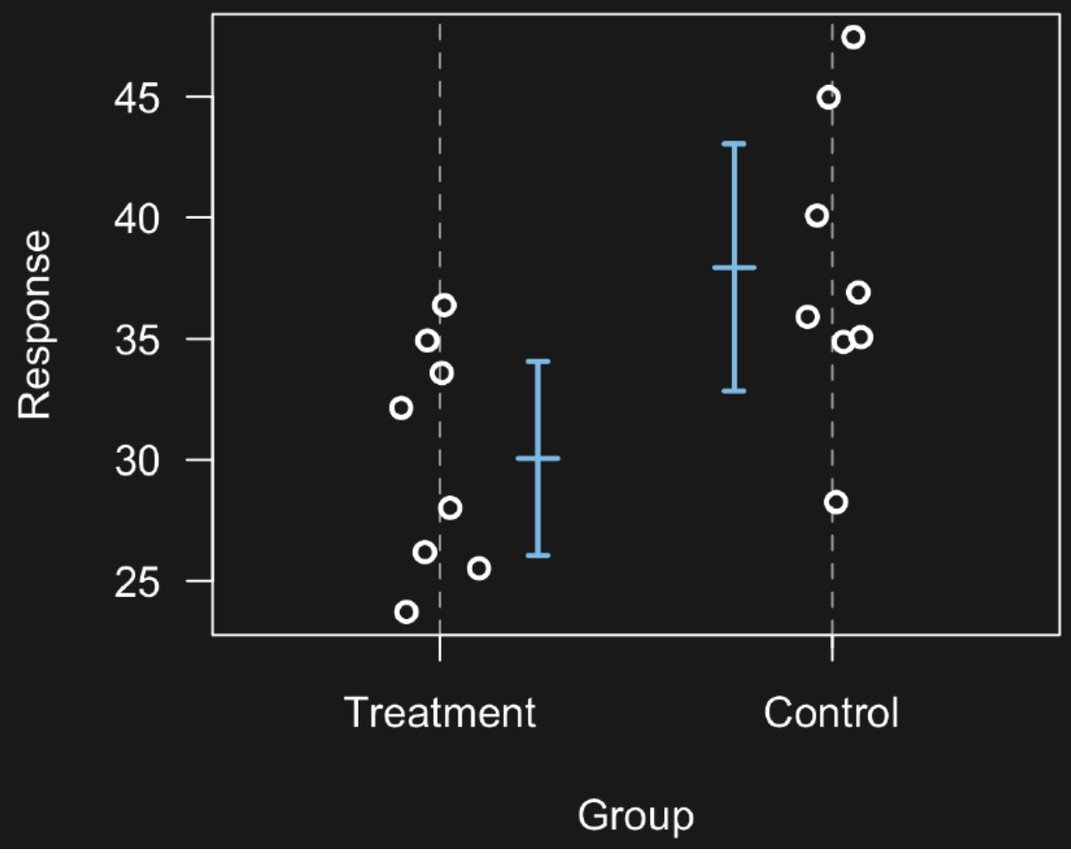


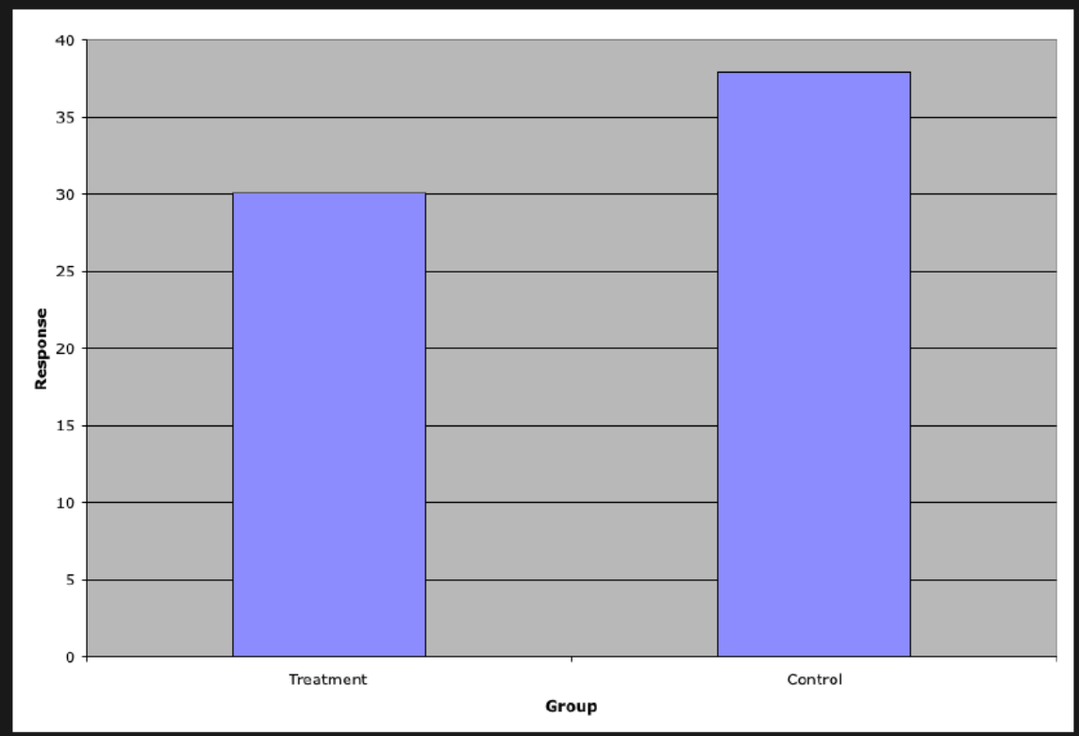


- No esconder los datos
- Mostrar la dispersión
- Mostrar la incertidumbre

http://kbroman.org/pages/talks.html

¿Qué NO hacer?

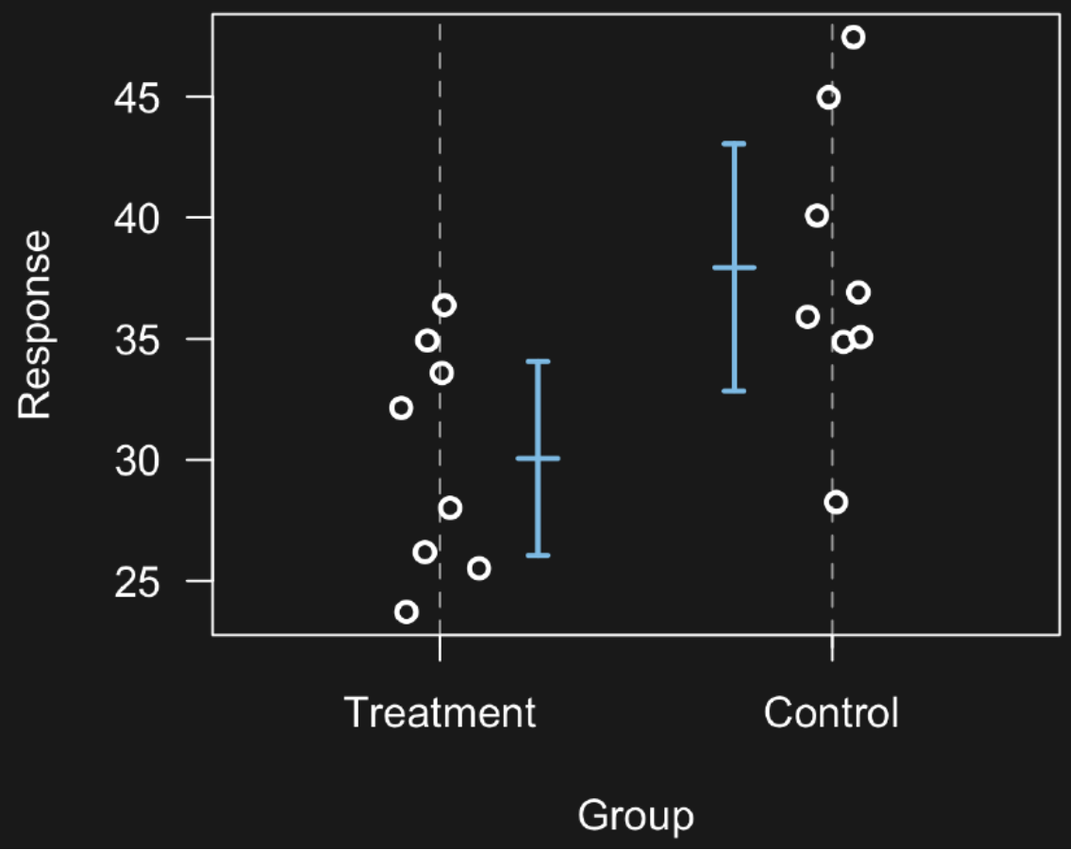


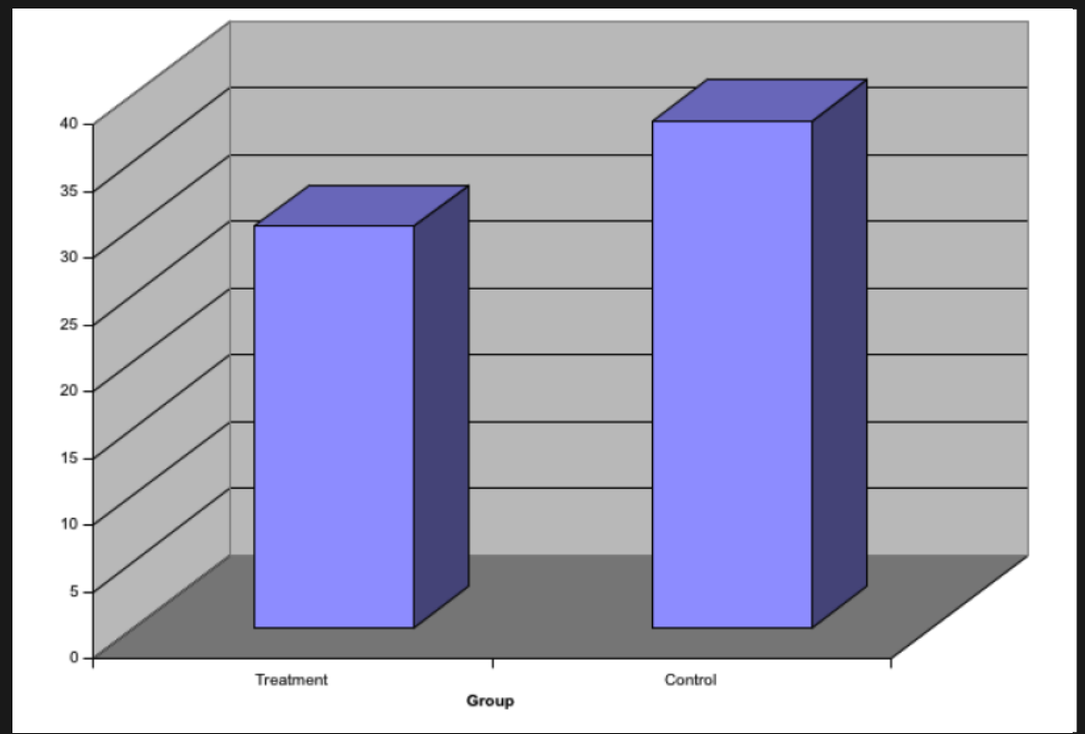


- No esconder los datos
- Mostrar la dispersión
- Mostrar la incertidumbre

http://kbroman.org/pages/talks.html

¿Qué NO hacer?

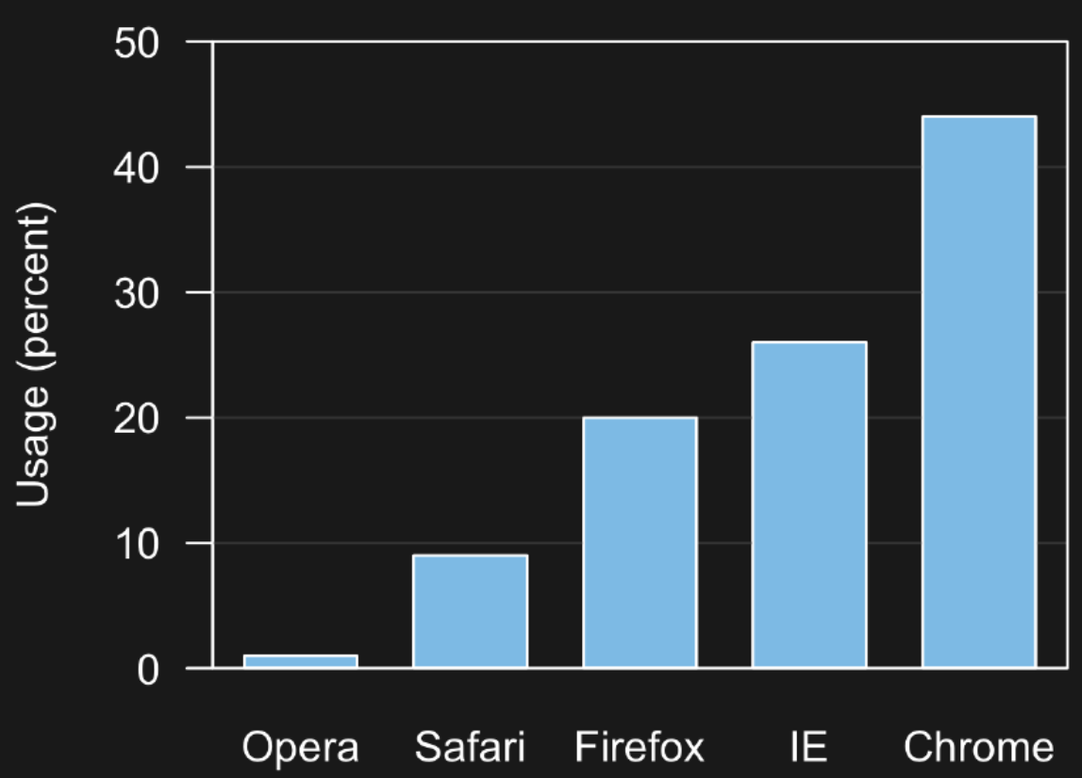


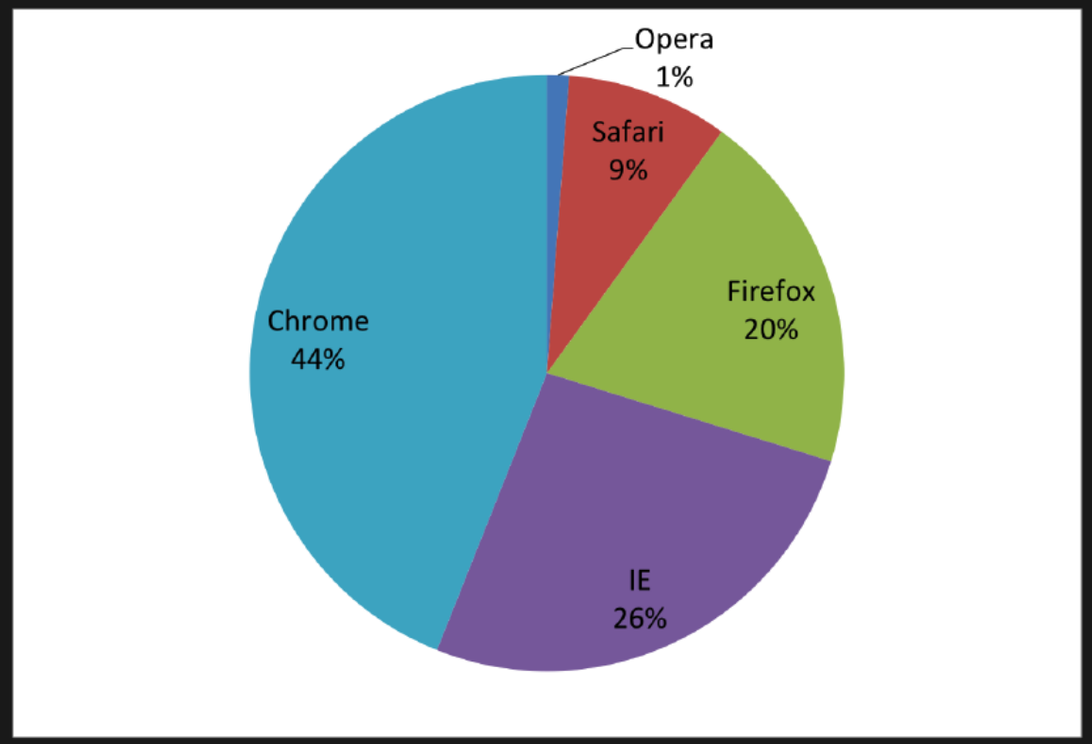


- No esconder los datos
- Mostrar la dispersión
- Mostrar la incertidumbre

http://kbroman.org/pages/talks.html

¿Qué NO hacer?

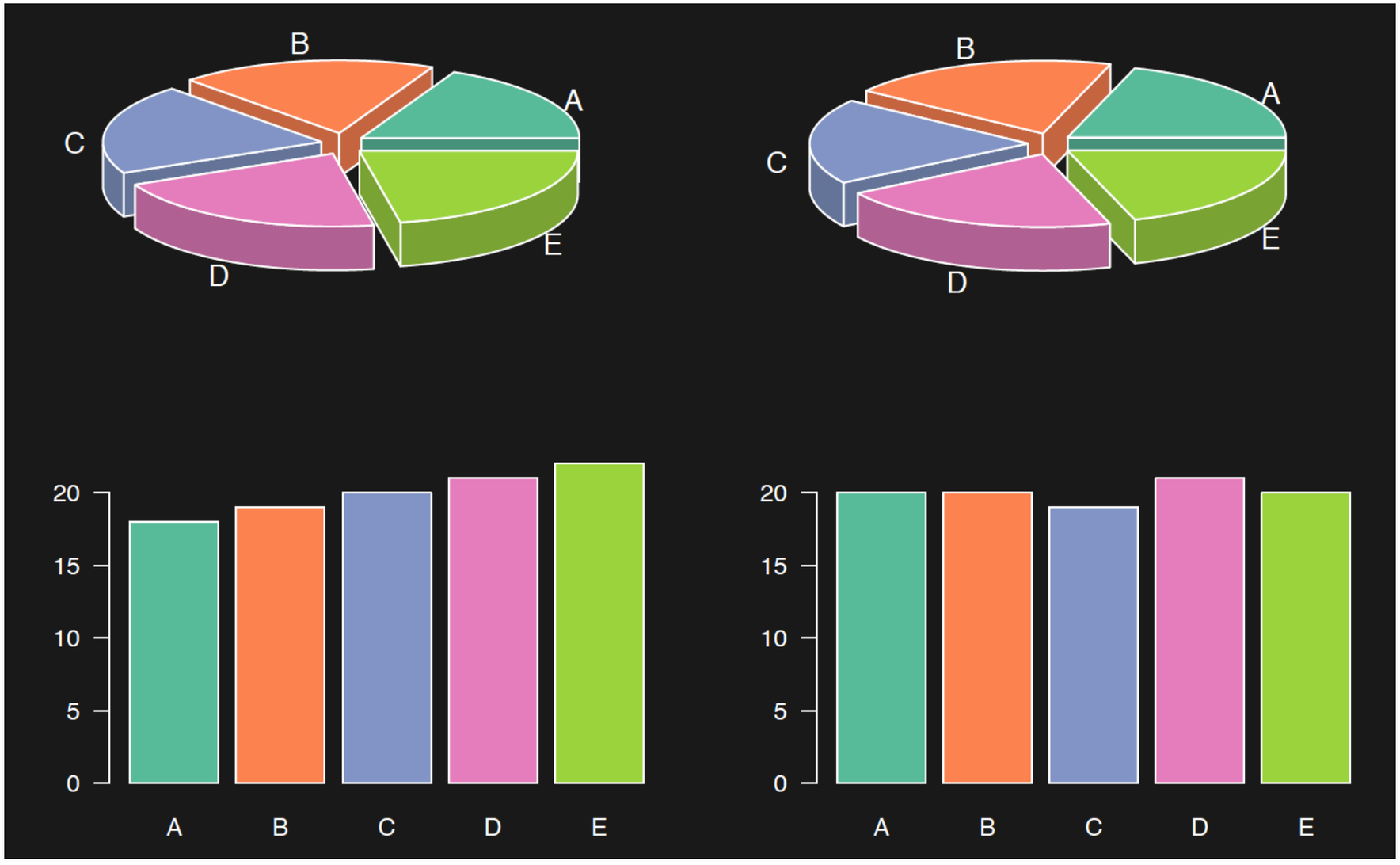




- Histogramas sobre gráficos de torta
- Apto para datos continuos, para mostrar distribución

http://kbroman.org/pages/talks.html

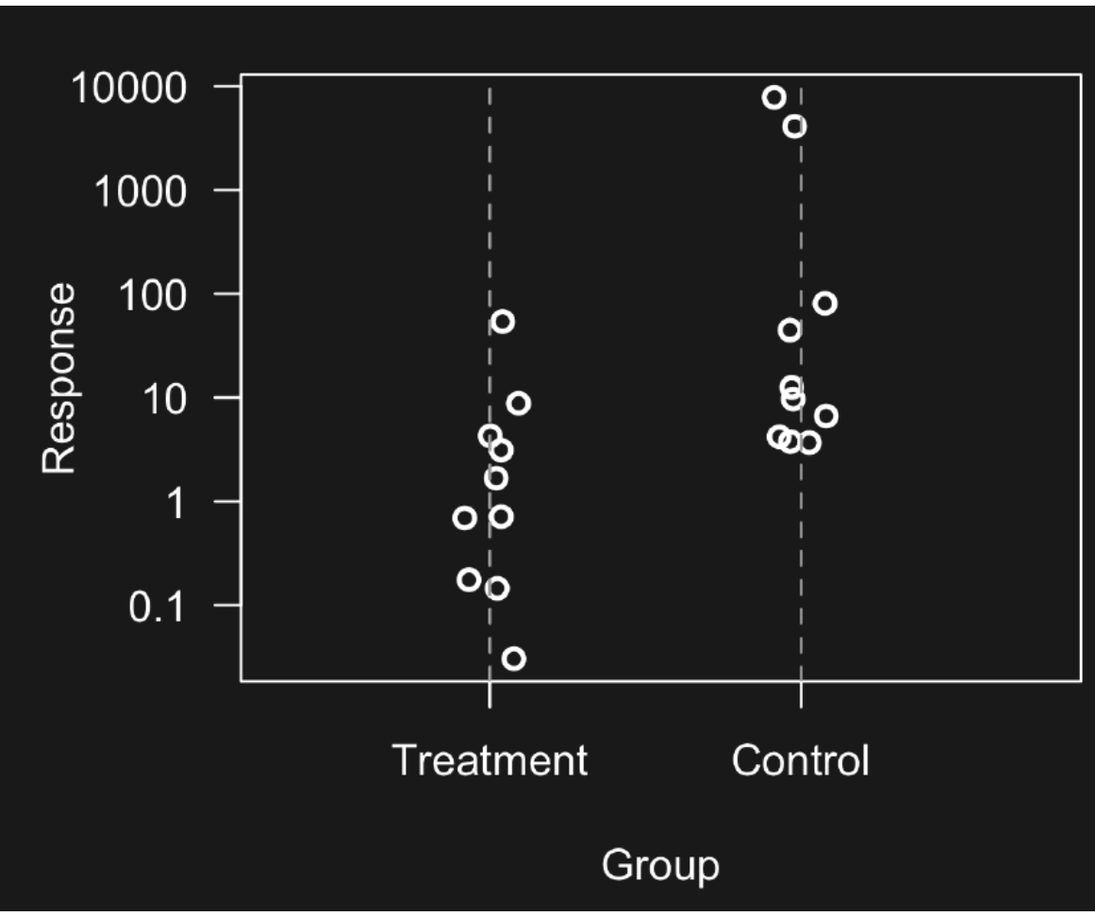
¿Qué NO hacer?

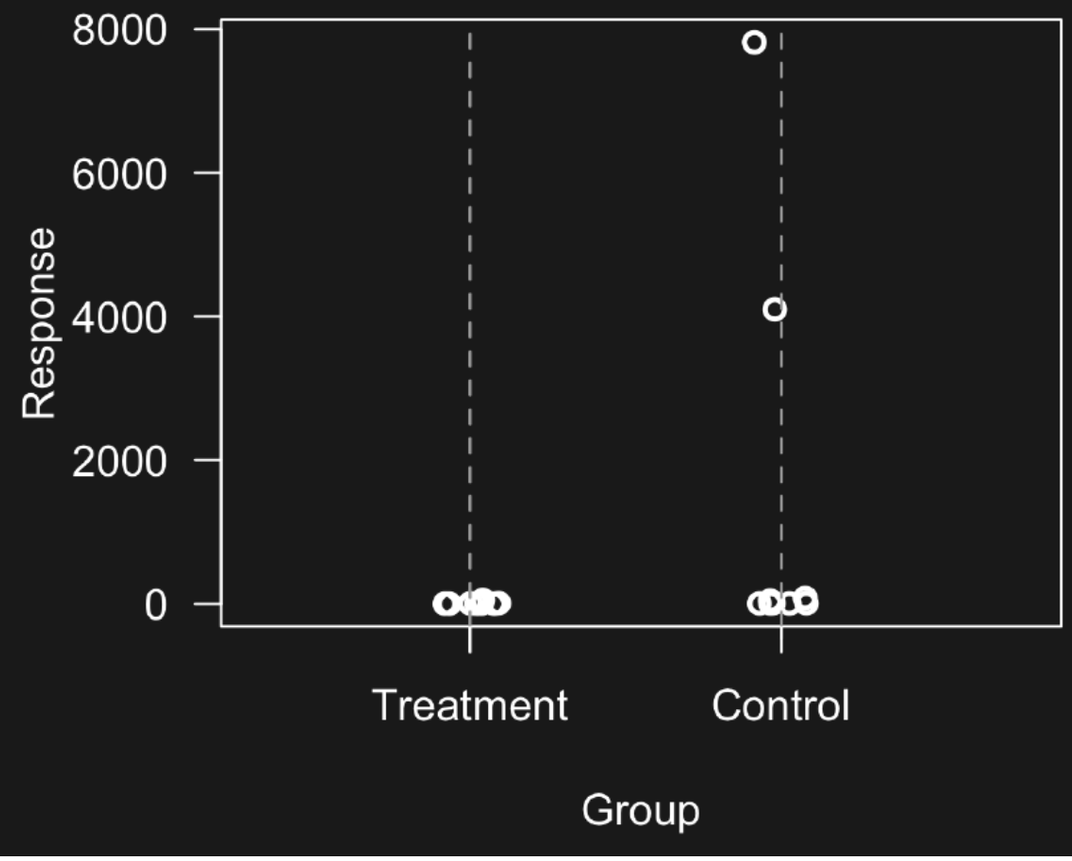


- Histogramas sobre gráficos de torta
- Apto para datos continuos, para mostrar distribución

http://kbroman.org/pages/talks.html

Considerar transformaciones de datos - Logaritmos

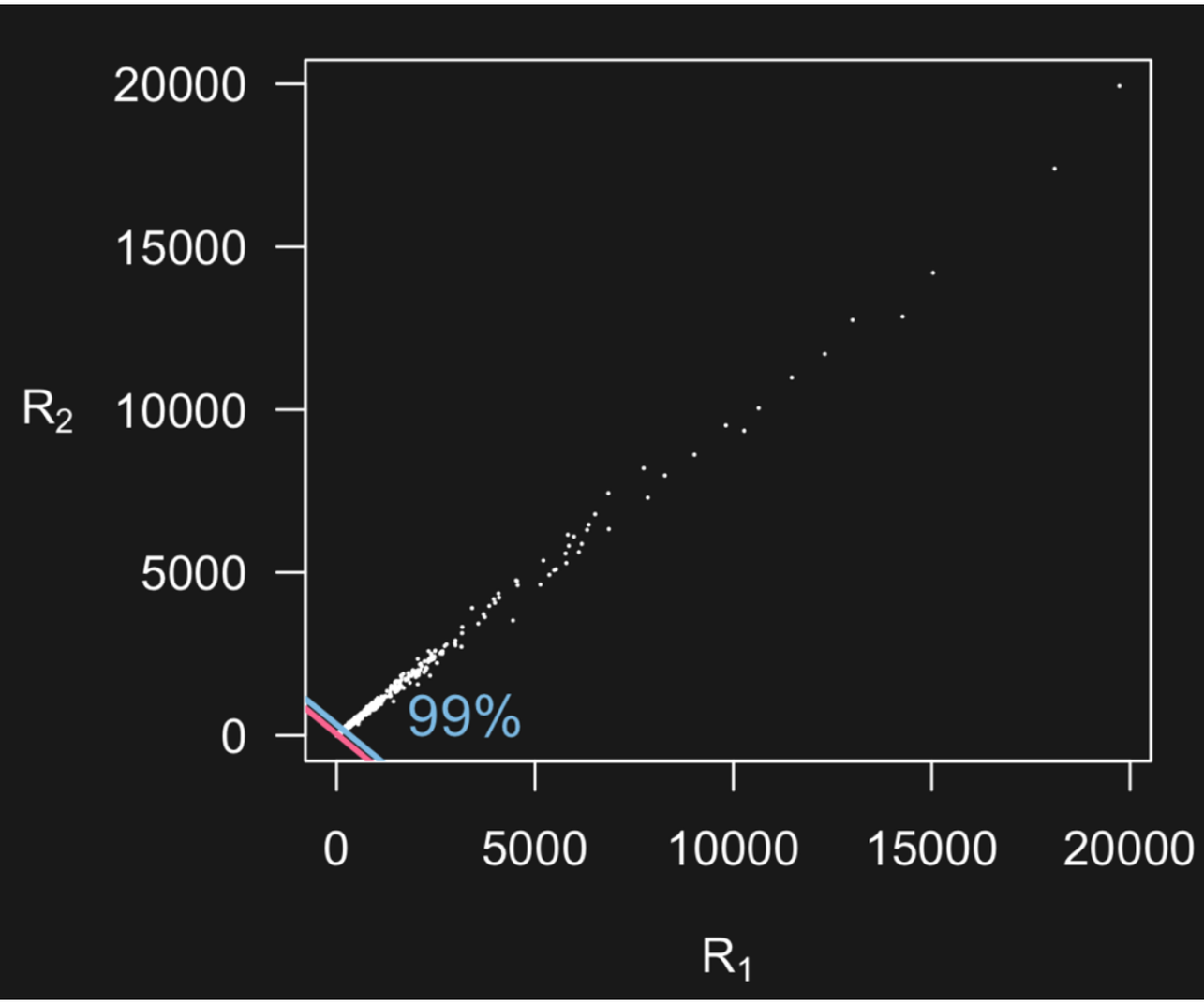


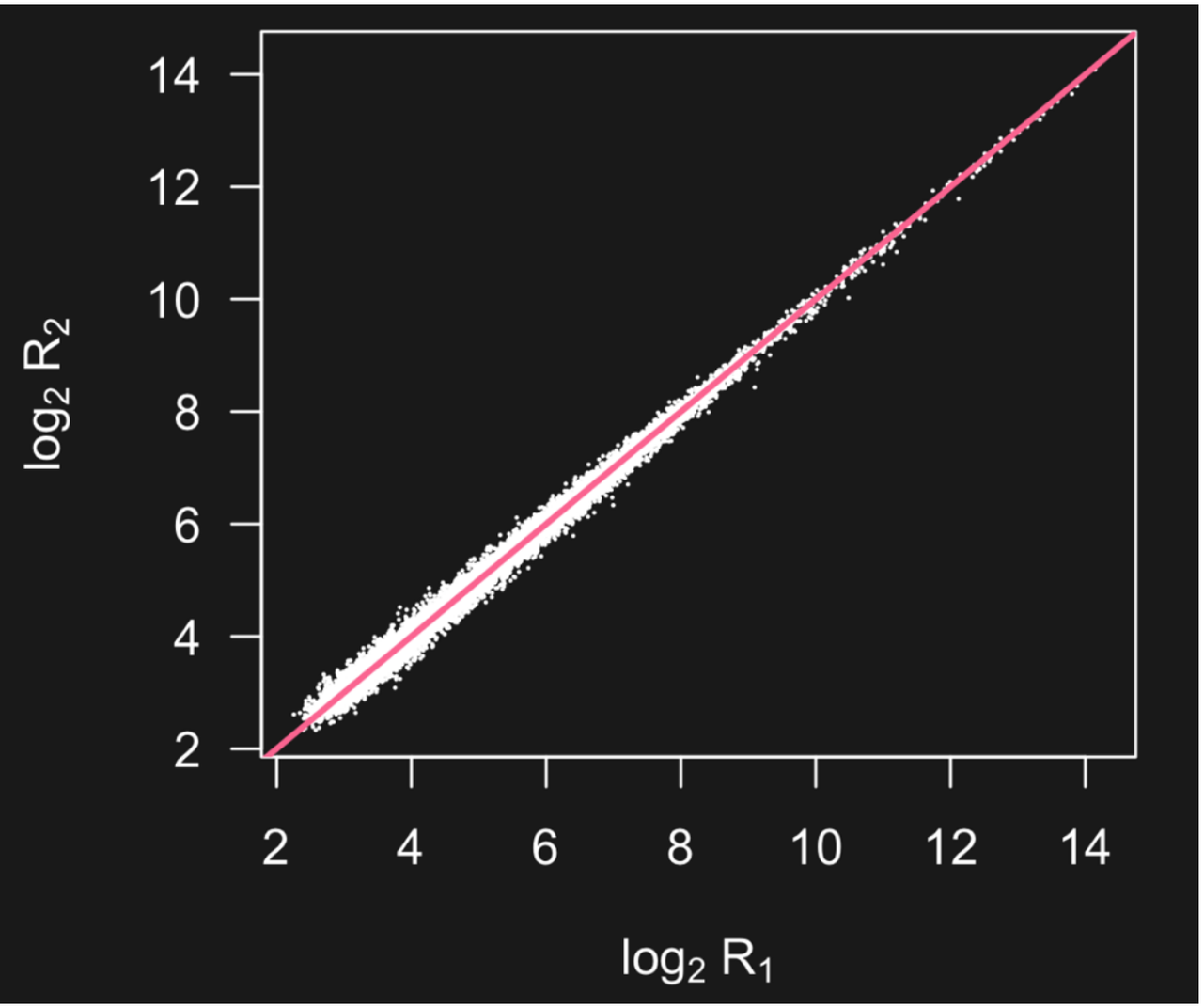


- ¿Qué gráfico muestra mejor los datos?

http://kbroman.org/pages/talks.html

Considerar transformaciones de datos - Logaritmos

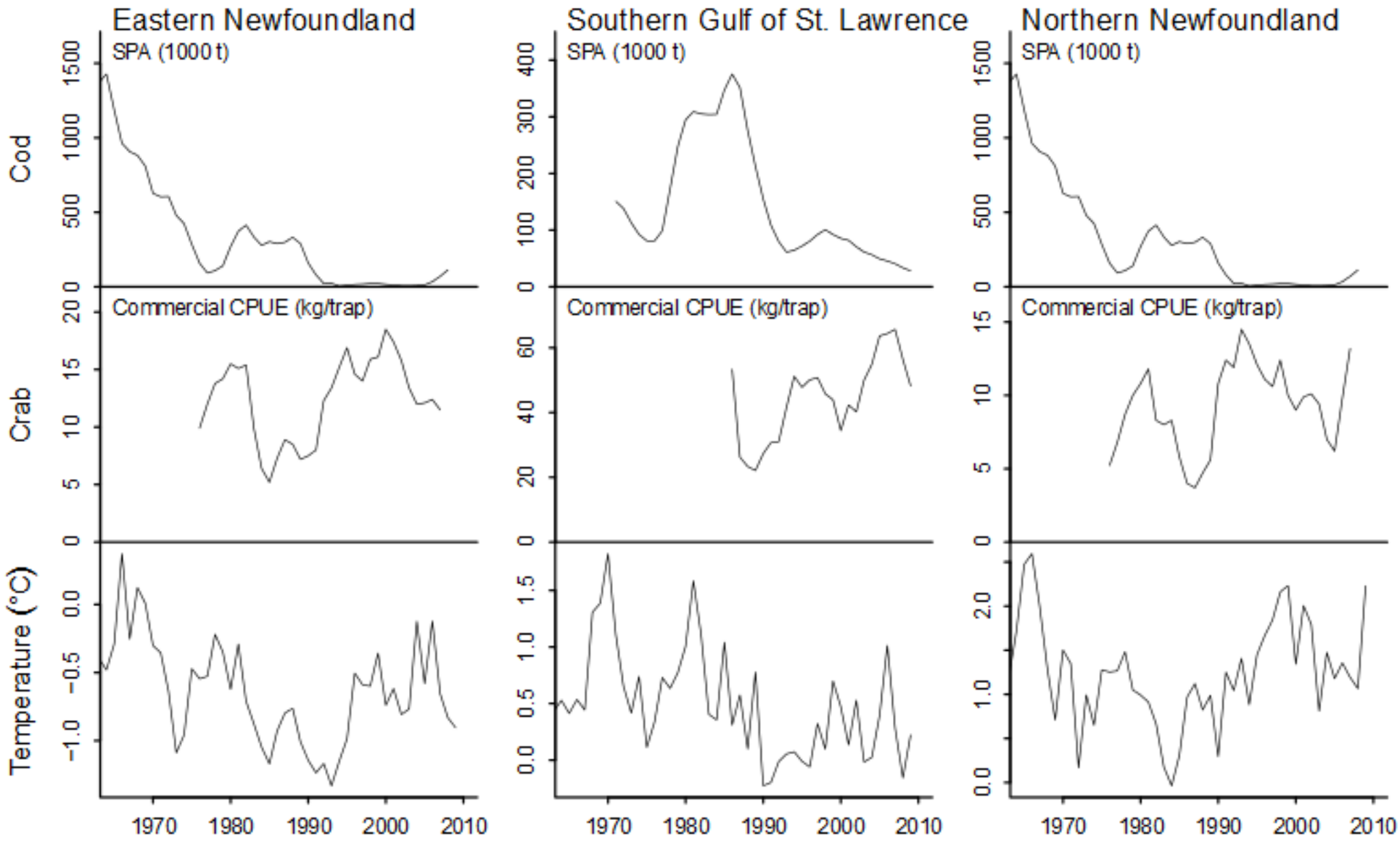




- ¿Qué gráfico muestra mejor los datos?

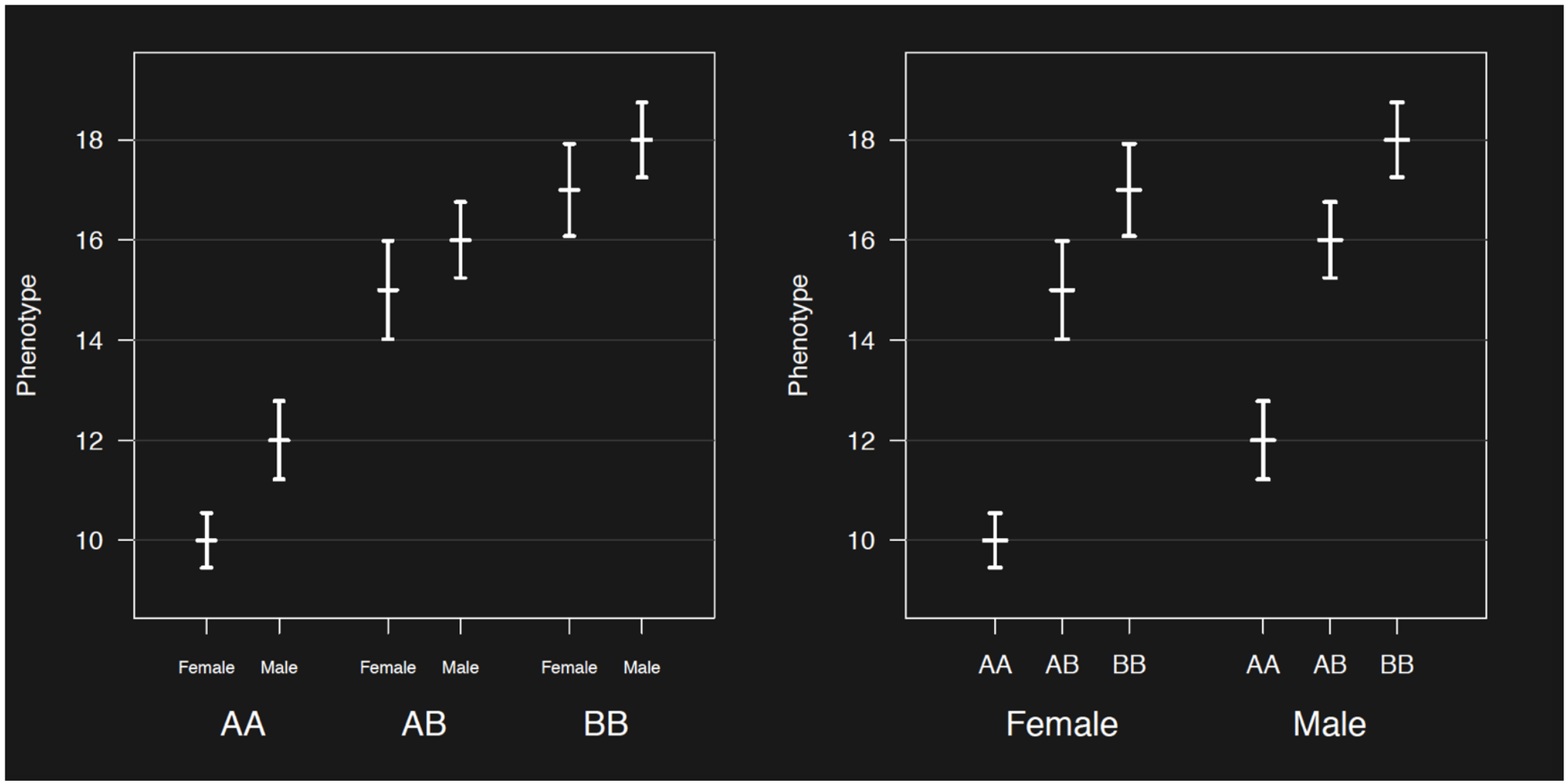
http://kbroman.org/pages/talks.html

¿Qué NO hacer?



- Las escalas son distintas
- Es difícil hacer comparaciones

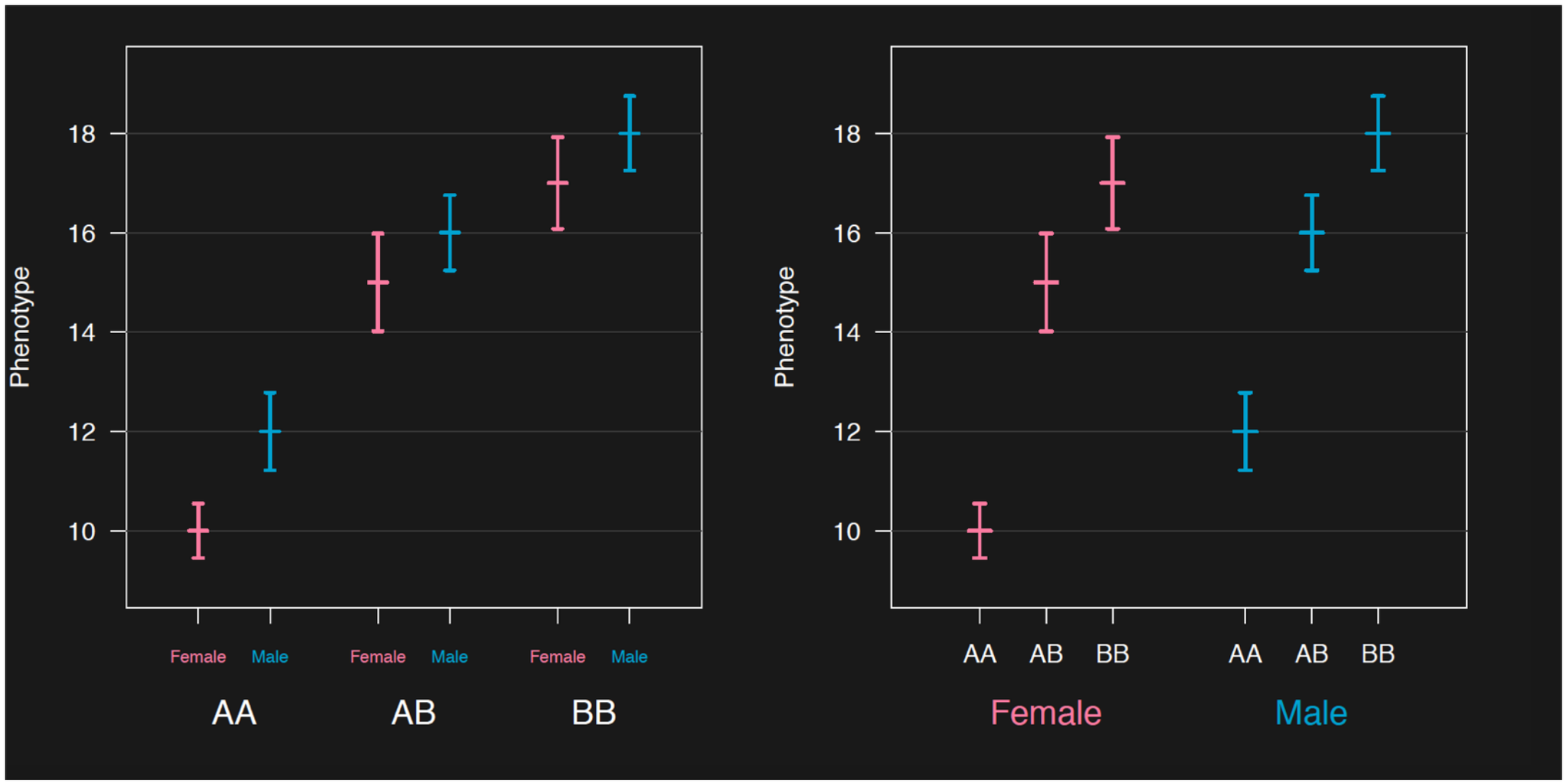
Gráficos para comparar deben ir lado a lado



- La escala debe ser igual para facilitar la comparación
- Agregar color para distinguir categorías de mejor forma

http://kbroman.org/pages/talks.html

Gráficos para comparar deben ir lado a lado



- La escala debe ser igual para facilitar la comparación
- Agregar color para distinguir categorías de mejor forma

http://kbroman.org/pages/talks.html

Usar colores de manera consistente entre figuras y paneles

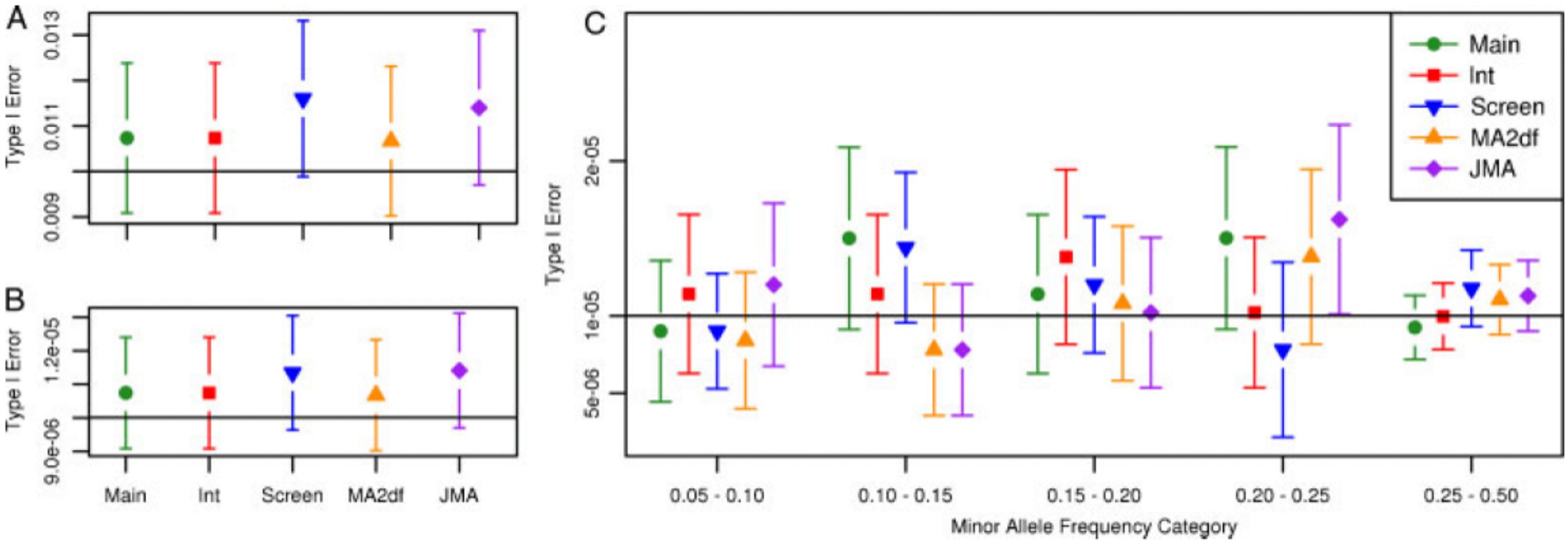
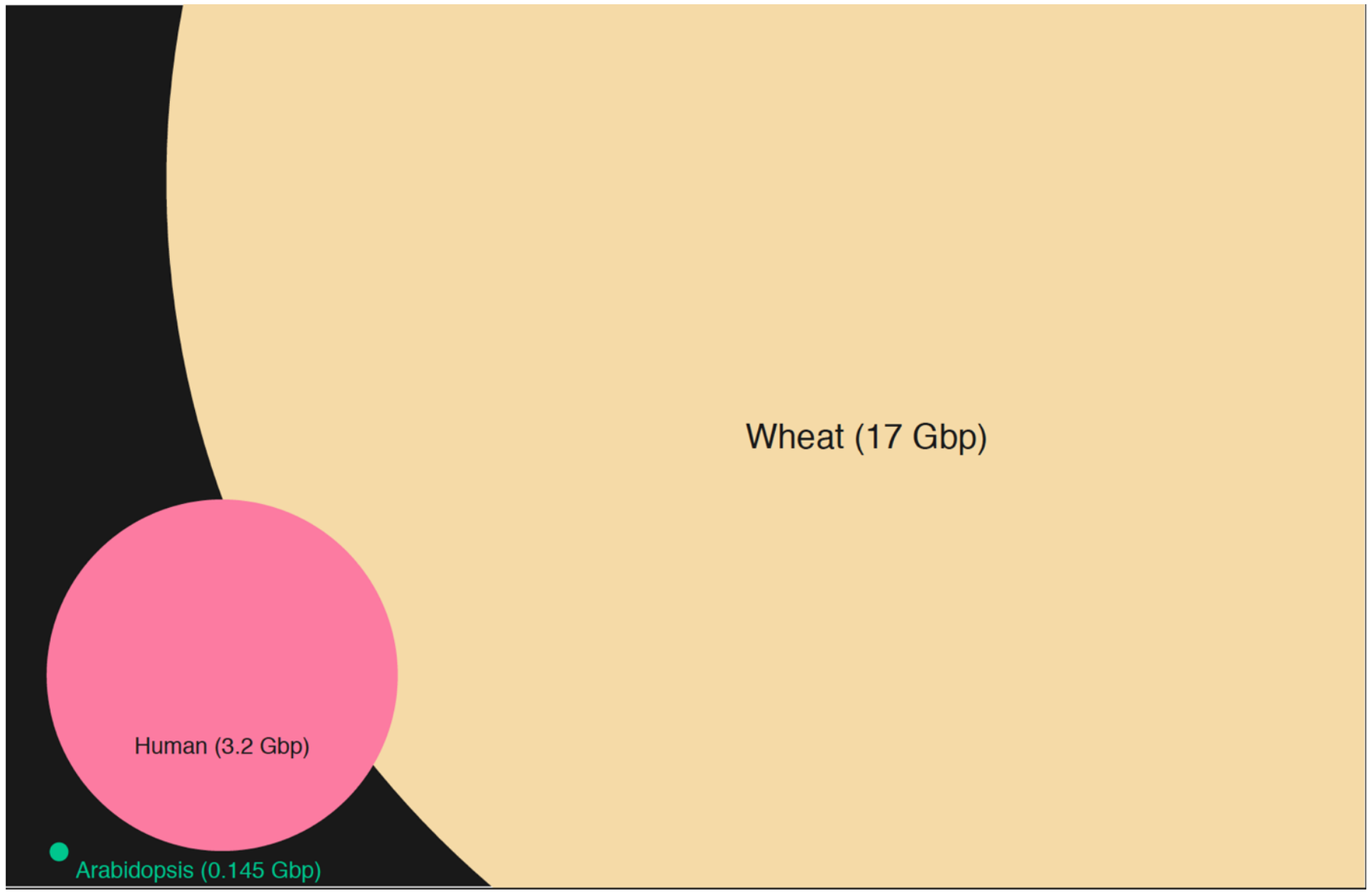


Fig. 1. (A) Empirical type I error rate in 15,000 simulations of 999 null SNPs, empirical type I error rates for (B) 14,985,000 null SNPs and (C) 14,985,000 null SNPs broken into minor allele frequency categories. A horizontal line is drawn at the theoretical type I error rate in each plot.

Manning AK, LaValley M, Liu CT, et al. Meta-Analysis of Gene-Environment Interaction: Joint Estimation of SNP and SNP x Environment Regression Coefficients. Genet Epidemiol 2011, 35(1):11-8

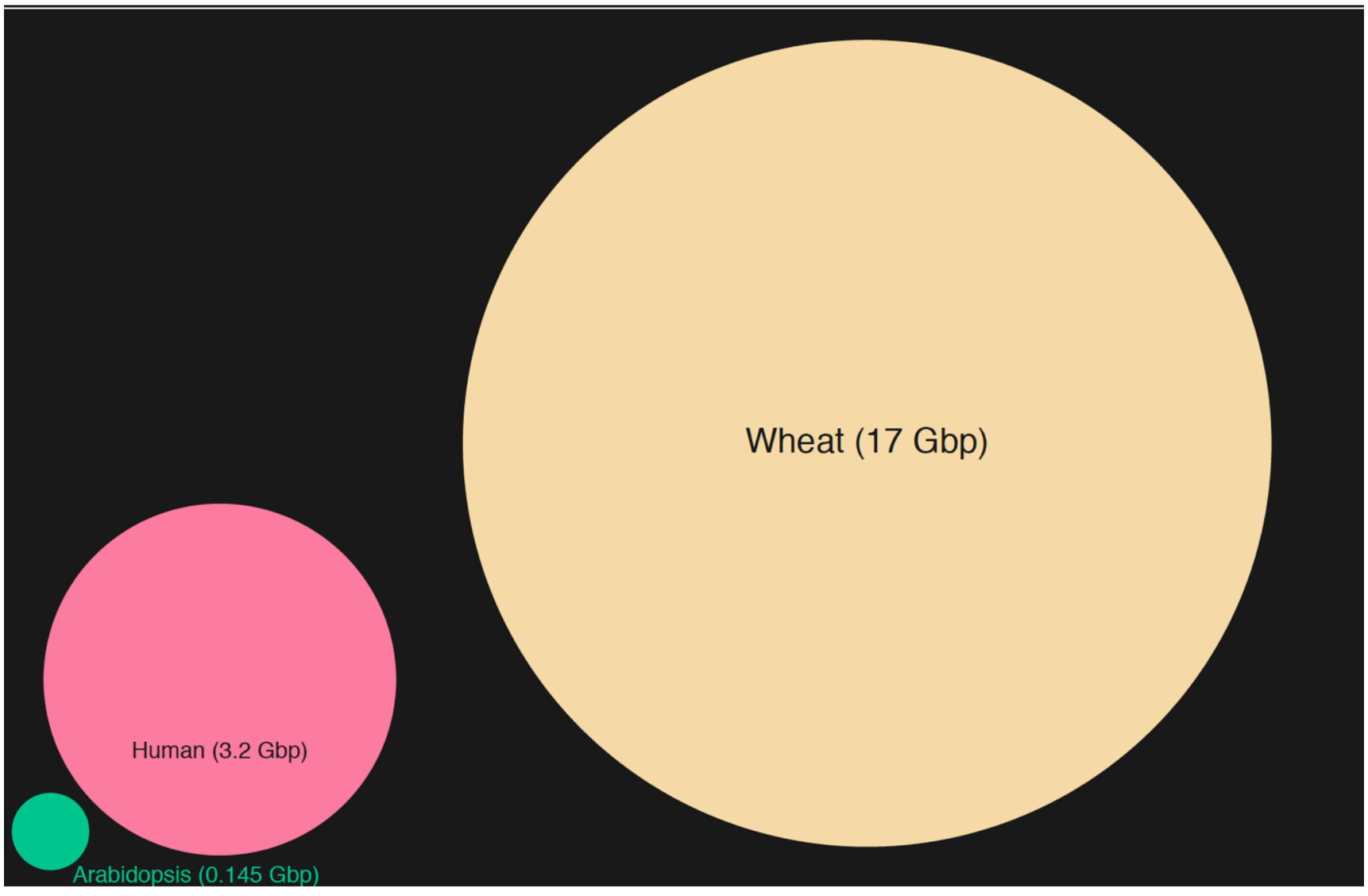
Círculos o polígonos deben ser proporcionales al área



- proporcional al radio

http://kbroman.org/pages/talks.html

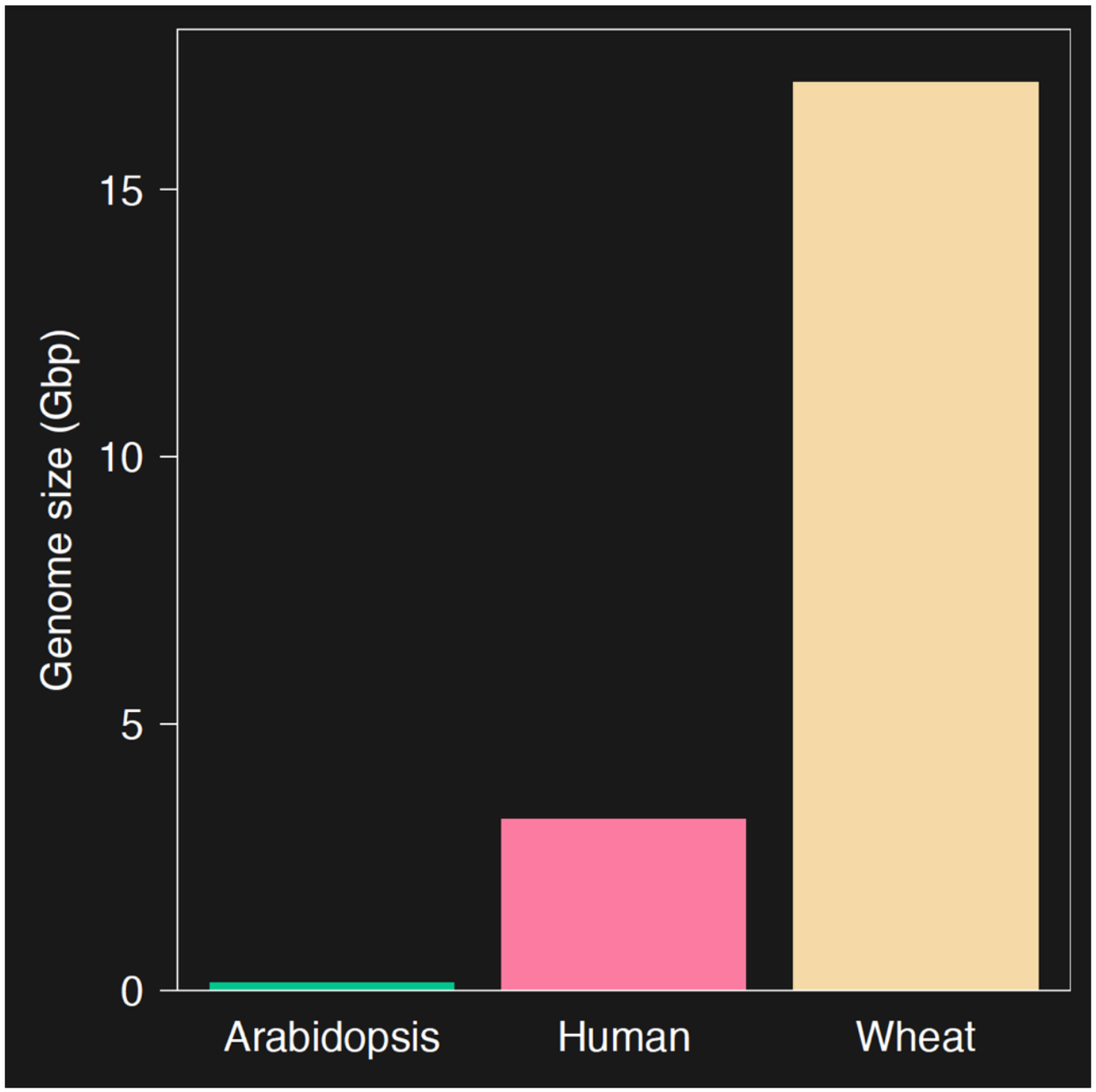
Círculos o polígonos deben ser proporcionales al área



- proporcional al área

http://kbroman.org/pages/talks.html

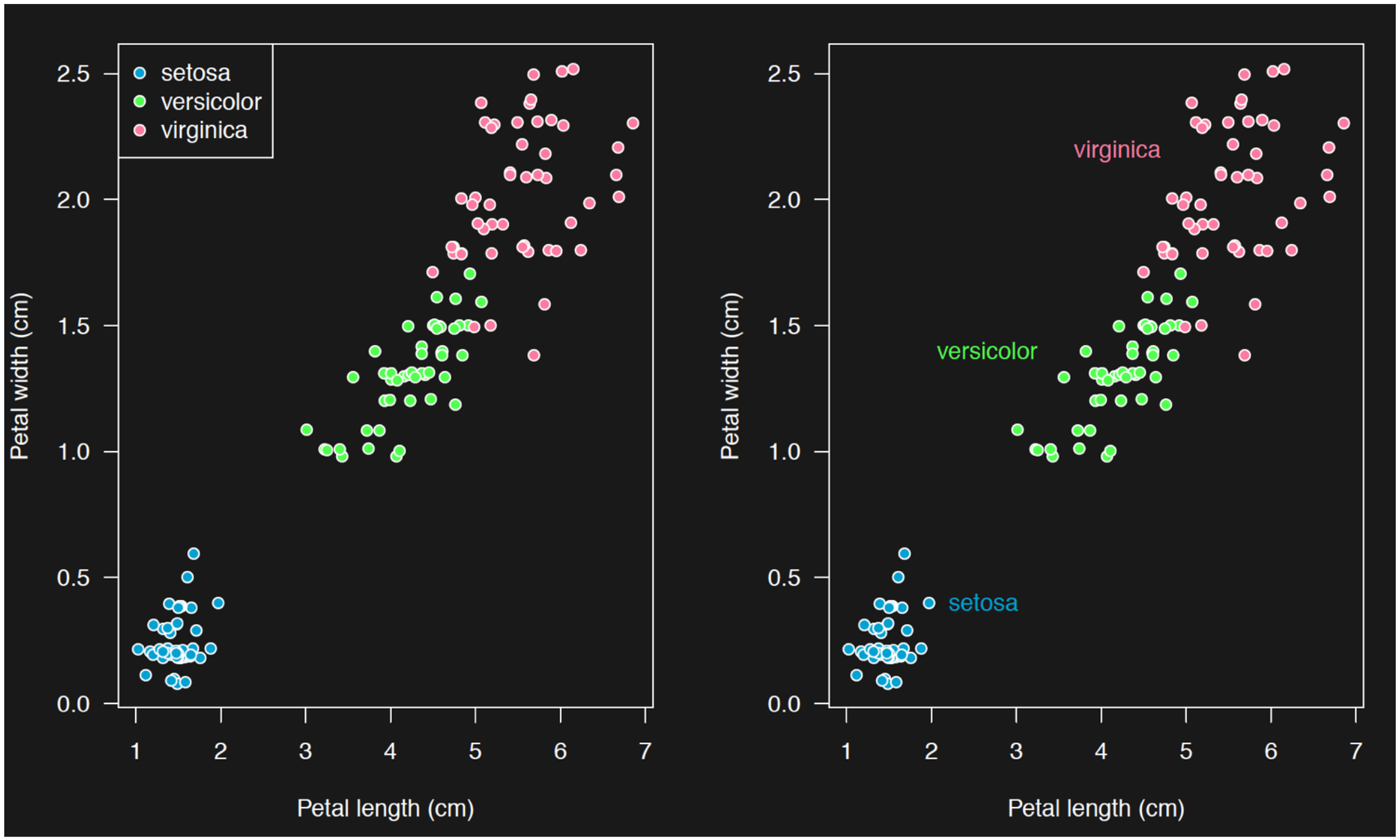
…O mejor usar barras



- proporcional a la longitud

http://kbroman.org/pages/talks.html

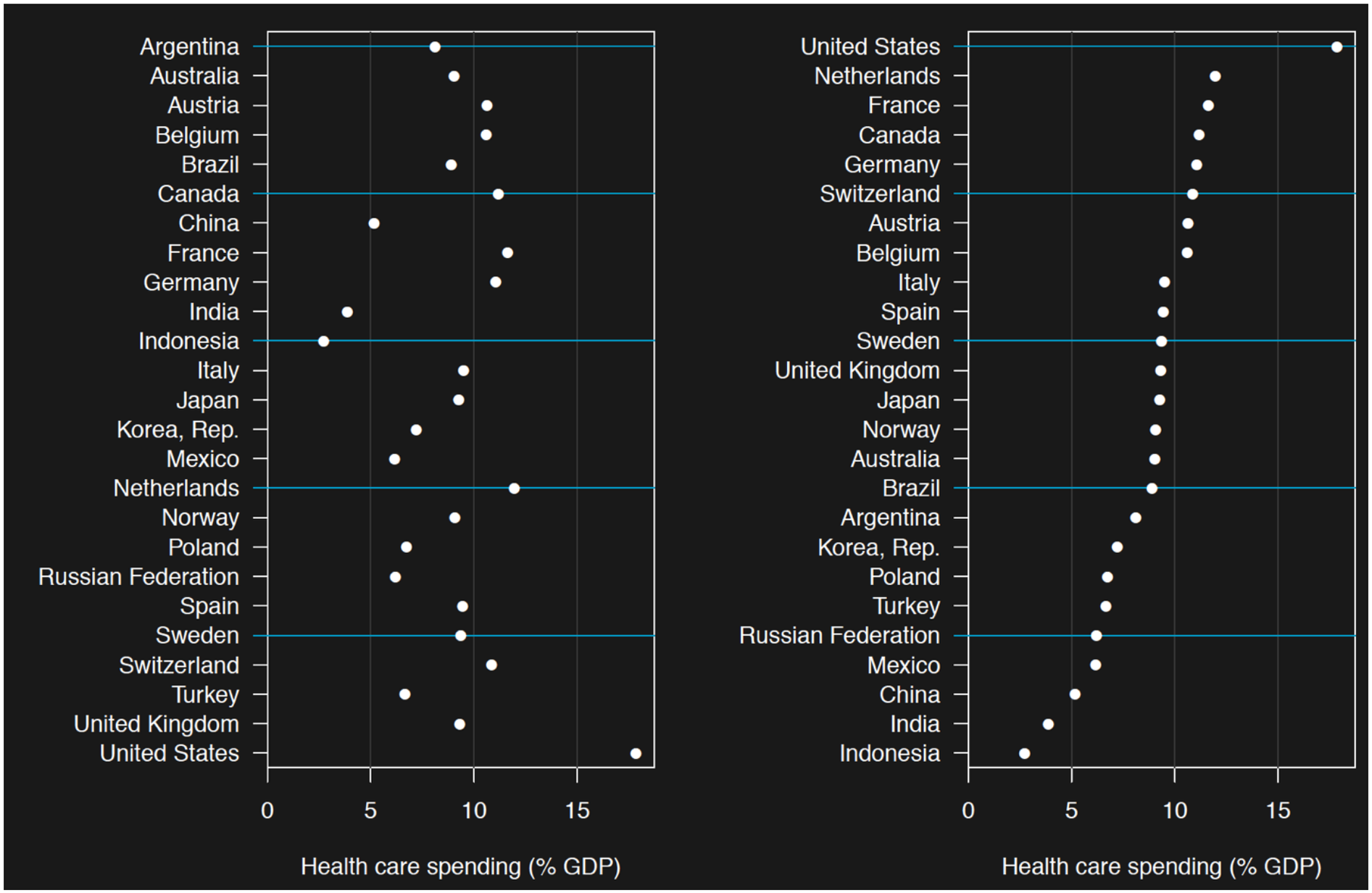
Prefieran etiquetas en vez de leyendas



- Etiquetas no quitan la atención del lector de los datos

http://kbroman.org/pages/talks.html

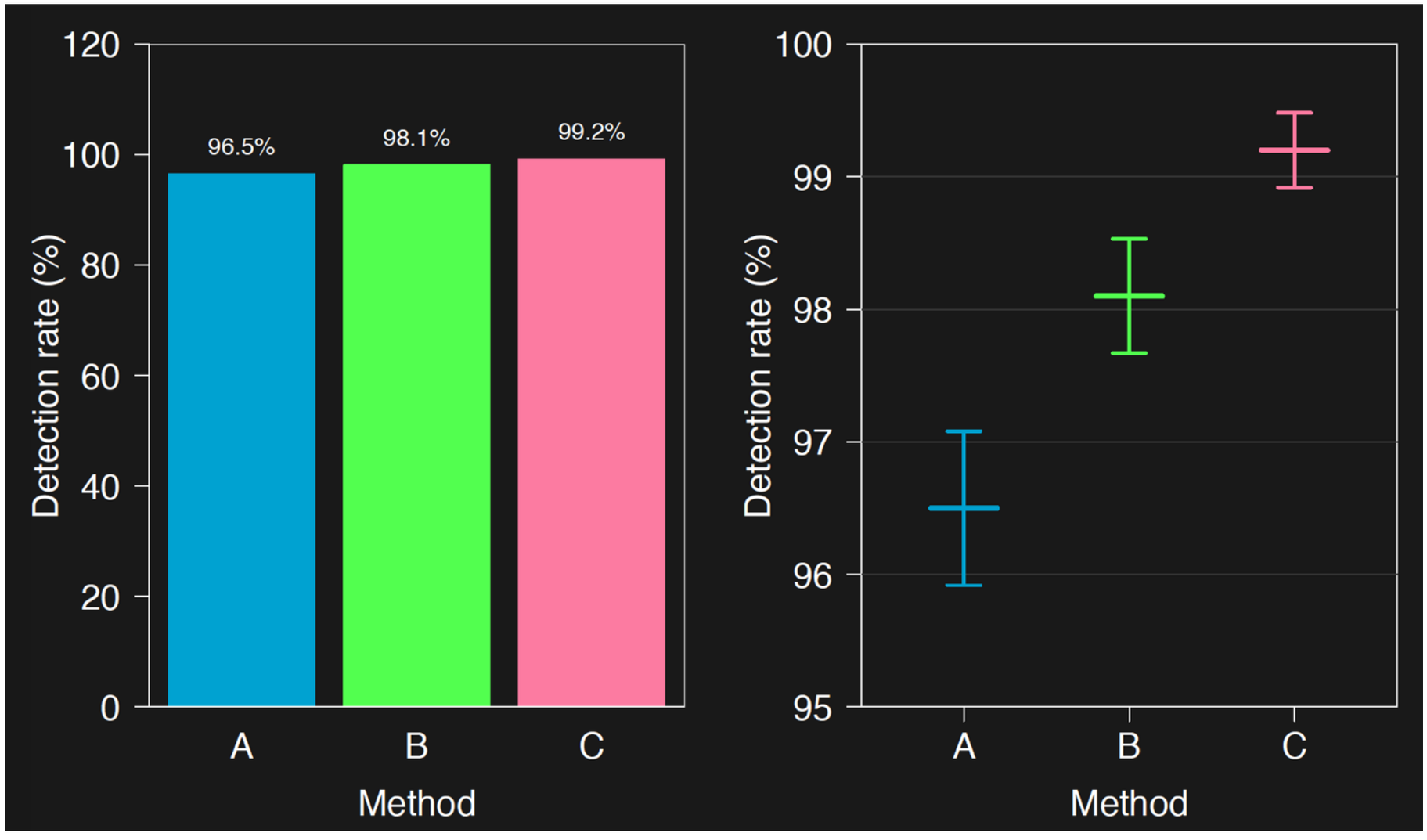
Nunca ordenar datos alfabeticamente



- Se pierde la estructura de los datos
- De mayor a menor las comparaciones son más eficientes

http://kbroman.org/pages/talks.html

A veces no es necesario incluir una escala completa



- Hacer énfasis en las diferencias

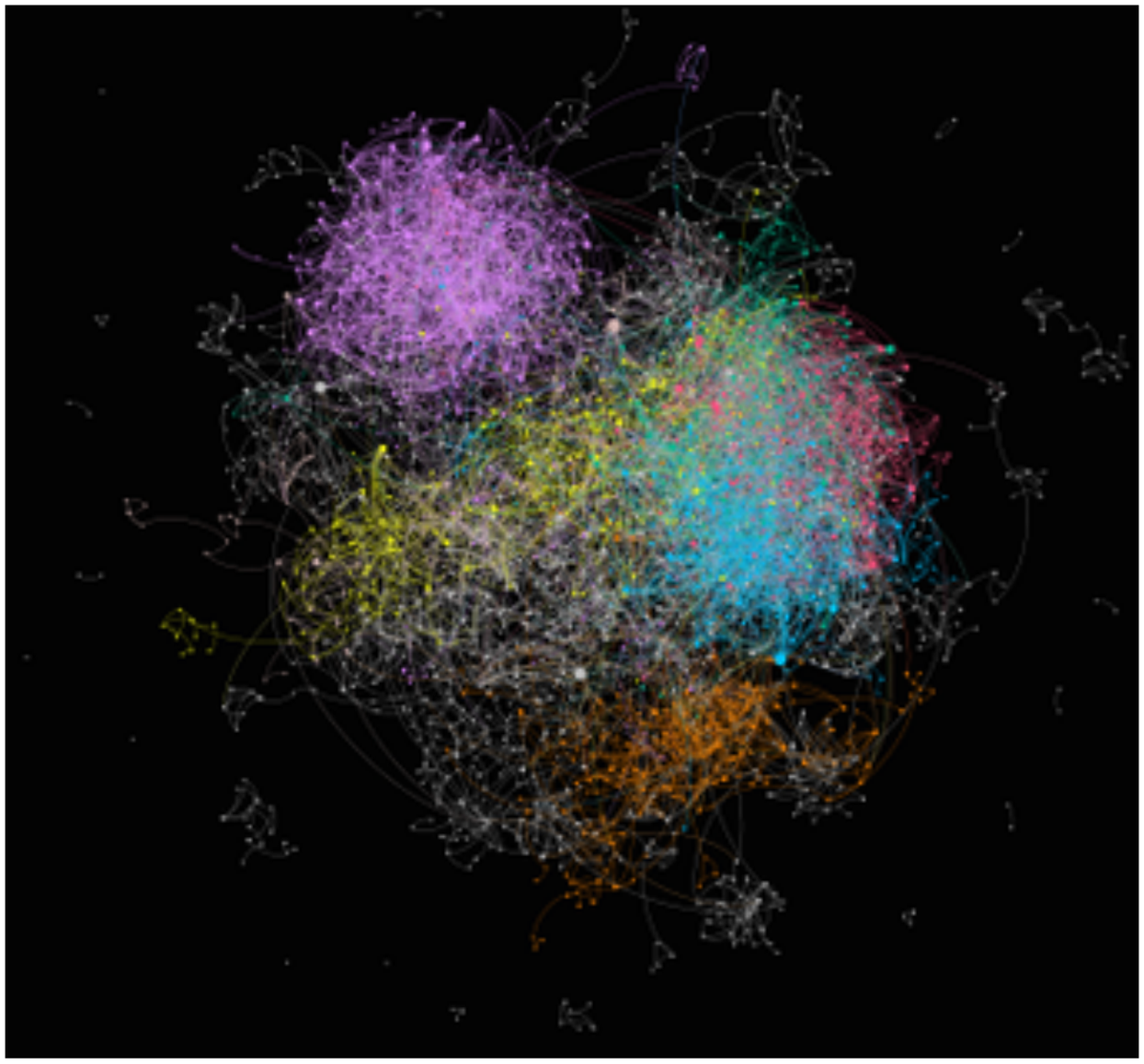
http://kbroman.org/pages/talks.html

Los 10 gráficos más malos

- https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/

El ridiculograma

- https://twitter.com/hashtag/ridiculogram



Preparen sus computadores

- Instalar R y R Studio
- Realizar el tutorial: http://www.datacarpentry.org/R-ecology-lesson/ de la lección 1 a la 6
- Averiguar lo que es un Tufte Handout
- Instalar el paquete de R “ggplot2”