

# Importancia de la Visualización Estadística

Instituto de Estadística (IESTA)  
Natalia da Silva  
@pacocuak

23 de noviembre de 2017

# De que voy a hablar

- Porqué y para qué visualizar
- Visualización estadística
- Ideas para una visualización efectiva

# Importancia de la Visualización

"The greatest value of a picture is when it forces us to notice what we never expected to see." Tukey (1977)

- Los cálculos numéricos son exactos pero los gráficos son toscos o aproximados
- Para cada tipo de datos estadísticos hay un conjunto de cálculos que constituyen un análisis estadístico correcto
- Hacer cálculos intrincados es virtuoso mientras que mirar los datos es hacer trampa, Anscombe (1973)

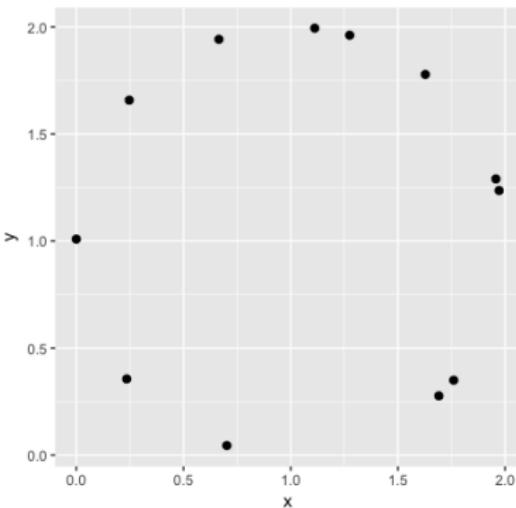
La visualización juega un rol importante en todas las etapas del análisis estadístico.

- **Exploración:** Encontrar patrones generales y específicos en los datos.
- **Modelado:** Chequear supuestos sobre los datos antes de modelar. Se puede hacer inferencia gráfica (nueva linea).
- **Diagnóstico:** Visualizar el modelo en el espacio de los datos ó los datos en el espacio del modelo.

¿Cuál es la relación entre x e y?

x	y
1.972	1.236
1.112	1.994
0	1.009
0.665	1.942
0.235	0.356
0.247	1.658
1.275	1.961
0.702	0.045
1.76	0.35
1.691	0.277
1.628	1.778
1.957	1.29

# ¿Porqué usamos visualización?



## Statistical Computing and Graphics

Let's Practice What We Preach: Turning Tables into Graphs

Andrew GELMAN, Cristian PASARICA, and Rahul DODHIA

# ¿Porqué usamos visualización?

- Los gráficos proveen más información que los resúmenes numéricos
- Anscombe's quartet  
Anscombe (1973)

$$n = 11$$

$$\bar{x} = 9,0$$

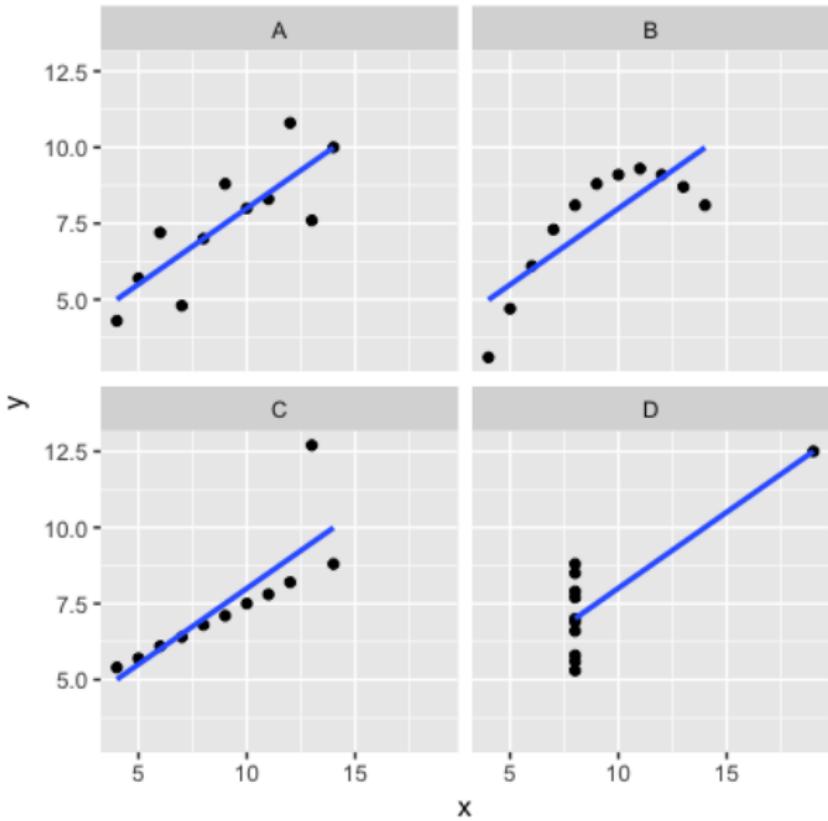
$$\bar{y} = 7,5$$

$$\hat{\beta}_1 = 0,5$$

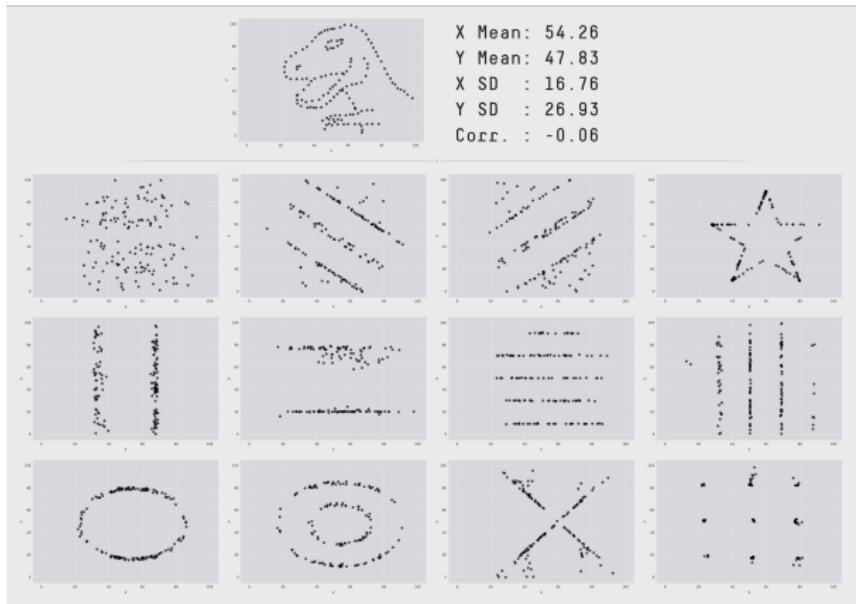
$$y = 3 + 0,5x$$

$$R^2 = 0,667$$

....



# ¿Porqué usamos visualización?



<https://github.com/stephlocke/datasauRus>

Los resúmenes estadísticos pueden ser los mismos pero las distribuciones muy diferentes

# Visualizaciones efectivas?

- No todas las visualizaciones son igualmente efectivas
- Hay diferentes criterios para evaluar gráficos (Cleveland, Tufte, Car, Wainer, etc )
- Basado en el estudio de la percepción gráfica de Cleveland (Cleveland and McGuill, 1985), las mejores visualizaciones son aquellas que requieren el uso de la visión "pre-attentive" (instantáneo, sin aparente esfuerzo visual)

## Graphical Perception and Graphical Methods for Analyzing Scientific Data

William S. Cleveland and Robert McGill

Graphs provide powerful tools both for analyzing scientific data and for communicating quantitative information. The computer graphics revolution, which began in the 1960's and has intensified during the past several years, stimulated the invention of graphical meth-

mation from graphs; theory and experimental data are then used to order the tasks on the basis of accuracy. The ordering has an important application: data should be encoded so that the visual decoding involves tasks as high in the ordering as possible, that is, tasks per-

al field that comes without apparent mental effort. We also perform cognitive tasks such as reading scale information, but much of the power of graphs—and what distinguishes them from tables—comes from the ability of our preattentive visual system to detect geometric patterns and assess magnitudes. We have examined preattentive processes rather than cognition.

We have studied the elementary graphical-perception tasks theoretically, borrowing ideas from the more general field of visual perception (7, 8), and experimentally by having subjects judge graphical elements (1, 5). The next two sections illustrate the methodology with a few examples.

### Study of Graphical Perception: Theory

A series of small, light-colored navigation icons typically found in LaTeX Beamer presentations, including symbols for back, forward, and search.

# ¿Cómo decodificamos un gráfico?

- Cuando miramos un gráfico la información es visualmente decodificada por el sistema visual de la persona
- Un método gráfico es efectivo solamente si la decodificación lo es
- Buenas visualizaciones son las que optimizan el sistema visual humano
- Si esto es verdad tendríamos que saber cómo el sistema humano decodifica un gráfico

Cleveland ordena la dificultad de los elementos gráficos basado en la percepción para estimar variables cuantitativas

- ① Posición a lo largo de una escala común
- ② Posición en escals no alineadas pero ideénticas
- ③ Longitud
- ④ Ángulo ó pendiente
- ⑤ Área
- ⑥ Volúmen, densidad ó saturación del color
- ⑦ Tono del color

- Cuál es la comparación más importante que quiero hacer cuando tengo variables cuantitativas
- Tenemos que codificar eso que quiero comparar usando los elementos gráficos en la tabla en orden ( 1. posiciones en la misma escala )

# Comparar millas por galón (mpg)

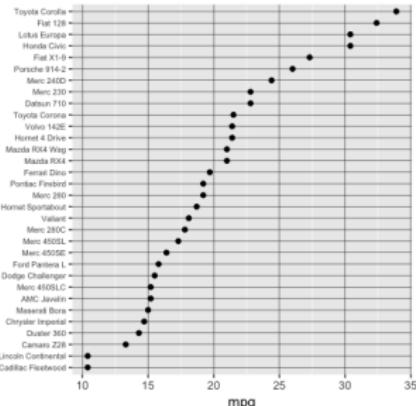
## 7) Color



## 3) Longitud

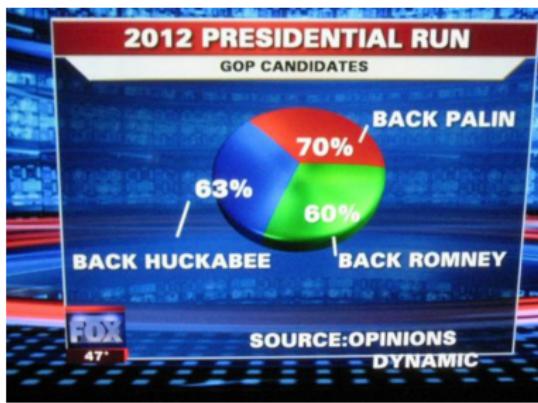
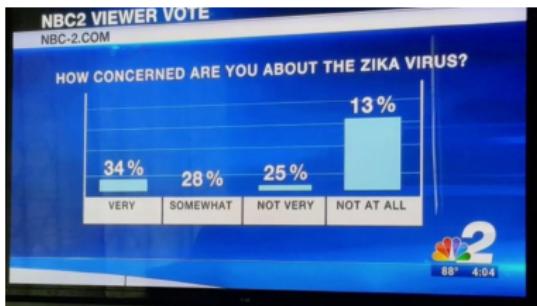


## 1) Posición, escala común



- Gráficos de torta son siempre un error!!!
- Pero ahora sabemos porqué
- Porque decodificar variables cuantitativas usando ángulos es más difícil que con otros elementos gráficos
- Siempre es preferible un gráfico de barras a uno de torta

# Visualización para comunicar adecuadamente



- ggplot2 (Wickham, 2016) es un paquete para producir gráficos estadísticos o de datos que a diferencia tiene una teoría que lo sustenta basada en Grammar of Graphics (Wilkinson, 2006)
- Aprender en base a una gramática gráfica te permite graficar cosas que sabés pero además crear nuevas y tal vez mejores gráficos para visualizar tus datos.

- Visualización es importante en todas las etapas del análisis estadístico
- Hacer visualización eficaz permite comunicar los resultados de forma sencilla
- Para estimar cantidades el elemento gráfico más eficaz es la posición a lo largo de una escala común
- Hacer visualización eficaz no es trivial, requiere conocimiento y trabajo
- Herramienta basada en gramática gráfica en R para producir gráficos de alta calidad, ggplot2

# Bibliografía

Anscombe, F. J. (1973), "Graphs in statistical analysis," *The American Statistician*.

Cleveland, William, S. and McGuill, R. (1985), "Graphical perception and graphical methods for analyzing scientific data," .

Tukey, J. W. (1977), "Exploratory Data Analysis," .

Wickham, H. (2016), *ggplot2: elegant graphics for data analysis*, Springer.

Wilkinson, L. (2006), *The grammar of graphics*, Springer Science & Business Media.