

## *Práctica gráfica en tesis de Actuaría (ITAM, 1990-2007)*

*Felipe González y grupo de Estadística Aplicada I, Otoño 2007  
14 de junio de 2013*

Los métodos gráficos son una herramienta poderosa en el análisis de datos. En contraste con otros métodos, sólo hasta recientemente han aparecido guías sólidas (aunque todavía parciales) de lo que constituye la buena práctica gráfica, y estas guías son relativamente poco conocidas.

Un punto de partida para la difusión de estas guías es el diagnóstico de la práctica actual, que varía mucho según el contexto (periódicos, industria, publicaciones científicas). En este trabajo buscamos producir estimaciones de varios indicadores básicos para las tesis de Actuaría en el periodo 1990-2007: qué tanto y qué tipos de gráficas se usan, y algunos indicadores de la calidad de la práctica gráfica.

### *Resultados principales y Recomendaciones*

Mostramos que las gráficas son utilizadas comunmente en las tesis de actuaría (en particular las de dispersión y las de barras), y que podrían mejorar en varios aspectos:

1. Reduciendo la basura gráfica en particular rejillas y rellenos,
2. Eliminando el uso de gráficas de pay, reduciendo el uso de gráficas de barras y sustituirlas por gráficas de puntos o tablas.
3. Usando más ampliamente gráficas de paneles múltiples (pequeños múltiples).

### *La mayor parte de las tesis de Actuaría requieren presentar datos en forma gráfica o tabular*

Diferentes tesis, por su naturaleza, requieren de más o menos representaciones de datos, y la elección es utilizar gráficas o tablas. El  $95\% \pm 3\%$  de las tesis continen al menos una gráfica o una tabla, y el  $80\% \pm 6\%$  contiene tanto gráficas como tablas.

Las dos gráficas de abajo sugieren que después de 2000 se ha incrementado el uso y presentación de datos en las tesis de Actuaría (total de gráficas y tablas presentadas), aún cuando la elección entre tablas y gráficas no ha cambiado (los estudiantes de actuaría deciden por tablas el  $64\% \pm 5\%$  de las veces. )

Estos resultados pueden indicar que aún cuando se usa más representaciones de datos, estas representaciones no han aumentado en tamaño o complejidad, lo que usualmente requiere de gráficas en vez

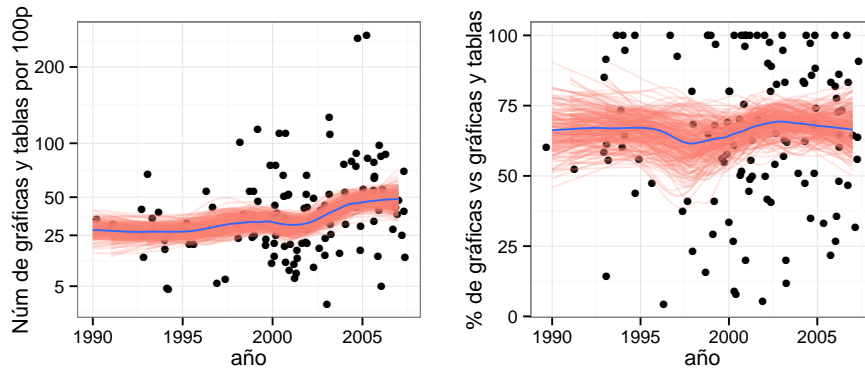


Figura 1: Gráficas de datos y tablas por año. Bandas construidas con 200 repeticiones bootstrap de un suavizamiento loess.

de tablas. Otra posible explicación es que las tablas cada vez se usan en situaciones donde gráficas serían más apropiadas.

Alrededor del  $84\% \pm 6$  de las tesis de actuaría contiene por lo menos alguna gráfica de datos. El número de gráficas por cada cien páginas es de  $18 \pm 3$ , y el 90% de las tesis con más gráficas utilizan gráficas de manera intensiva (más de  $[27, 50]$  gráficas por cada cien páginas. Abajo mostramos los cuantiles estimados, con algunas referencias más o menos conocidas.<sup>1</sup>

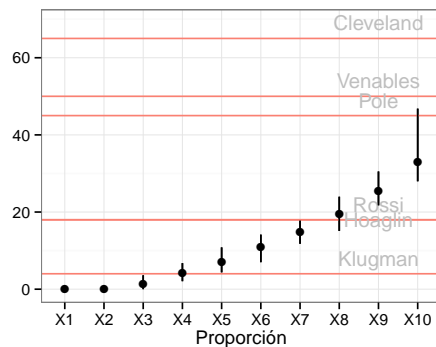


Figura 2: Deciles estimados de número de gráficas de datos.

<sup>1</sup> Cleveland, William S., *The Elements of Graphing Data*, Hobart Press, 1994.

Venables, W.N. y Ripley, B.D., *Modern Applied Statistics with S-PLUS*, Springer, 2002.

Pole A., West M. y Harrison, J., *Applied Bayesian Forecasting and Time Series Analysis*, Chapman & Hall, 1994.

Hoaglin, D.C., Mosteller, F. y Tukey, J.W., *Understanding Robust and Exploratory Data Analysis*, Wiley, 1983.

Rossi, P.E., Allenby, G. y McCulloch, R., *Bayesian Statistics and Marketing*, Wiley, 2005.

Klugman, S. A., *Loss Models: From Data to Decisions*, Wiley-Interscience, 2004.

Los tres casos que tienen densidad gráfica mayor al libro de Cleveland son interesantes: el mayor, con 123 gráficas por 100 páginas (IAC-2005-EAR, *Análisis comparativo del nivel de religiosidad en estudiantes católicos de dos universidades...*) contiene algunas gráficas de barras y muchas gráficas de pay decorativas acompañadas de tablas de datos (que es el estilo usual de la graficación de datos en la mercadotecnia). El segundo (IAC 2003.RRI) es un dato atípico por varias razones: se trata de una tesis de 31 páginas (sin contar bibliografía). Finalmente tenemos IAC-1999-FAN, *Sistema Bonus-Malus 200 para la conservación de la cartera*, que es atípica por el uso extensivo de histogramas (para representar distribuciones de pérdida).

### *Tipos de gráficas*

Las gráficas más comunes son las dispersión (que no son series de tiempo). En la siguiente tabla mostramos nuestras estimaciones para varios tipos de gráficas que fueron medidos para la población de tesis que tienen al menos una gráfica de datos.

	tipo	% de gráficas		% de tesis	
1	Dispersión	28	3	68	4
2	Barras	25	3	51	4
3	Serie de tiempo	24	4	52	4
4	Histograma	9	2	24	4
5	Circular	9	3	17	3
6	Caja y brazos	2	1	9	2
7	Mapas	1	1	4	2
8	Otras	0	0	2	1

Las gráficas de barras son probablemente sobreutilizadas, y podrían sustituirse con otro tipo o con tablas. El uso de pays es considerable.

Existen claras correlaciones entre la aparición de distintos tipos de gráficas dependiendo de la tesis: a grandes rasgos, las gráficas de barras y circulares tienden a aparecer juntas: son tesis que podrían parecer presentaciones de *marketing*. Por otra parte, tenemos también algunas tesis que utilizan gráficas más comunes en la estadística, como histogramas y gráficas de dispersión:

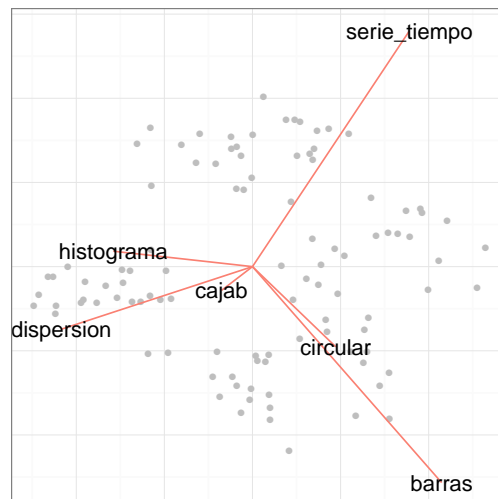


Figura 4: Biplot de primeras dos componentes principales para tabla binaria de tesis por tipo de gráficas

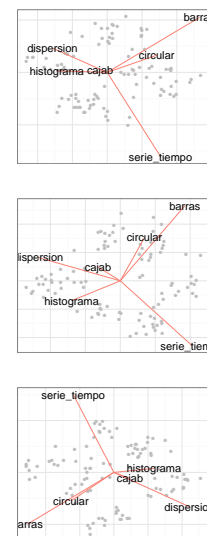


Figura 3: Replicaciones bootstrap de la gráfica de la izquierda

### Indicadores básicos de calidad gráfica

En nuestro estudio tenemos dos indicadores básicos de la calidad de cada gráfica:

- Contenido de basura gráfica (*chartjunk* de Tufte, como efectos 3d, *grid* o rejillas que ocultan o dificultan la percepción de los puntos de datos, marcos superfluos, etc.).
- Si la explicación de la gráfica está en la misma página de la gráfica o no. En el mejor trabajo gráfico texto, tablas, diagramas y gráficas se integran para explicar los datos.
- Uso de pequeños múltiplos, que consisten en la replicación de un mismo diseño y variando los datos en cada replicación.

Entre las tesis que contienen gráficas, el  $75\% \pm 7$  contiene algún tipo de basura gráfica, y el  $47\% \pm 7$  de las gráficas totales contiene algún tipo de basura. Estos números indican que las tesis no se dividen claramente en tesis sin basura y con basura.

Distintos tipos de basura gráfica están asociados a distintos tipos de gráficas:

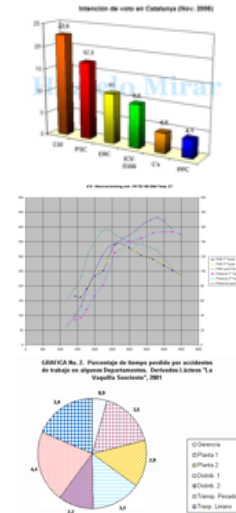


Figura 5: Tipos de basura gráfica: efecto 3d, rejilla, trama

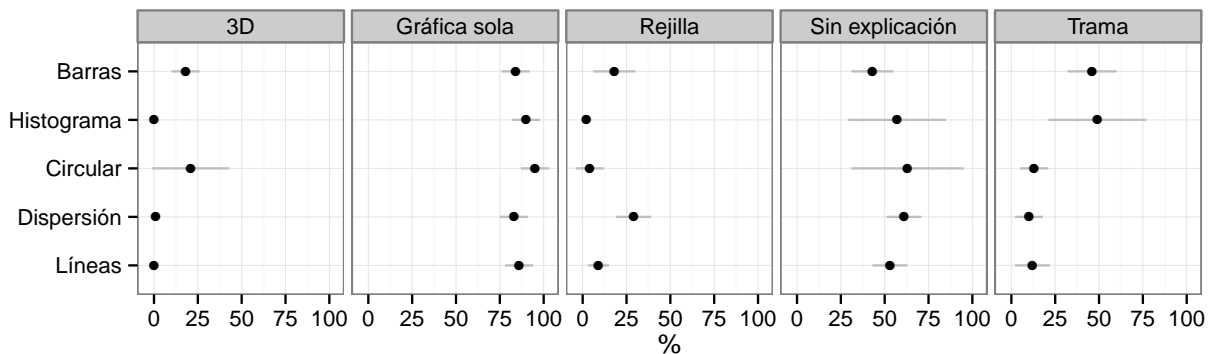


Figura 6: Tipos de basura gráfica registrados según tipo de gráfica. Los criterios de basura gráfica son subjetivos (una rejilla se considera basura si tiene un grosor y color similar a los datos graficados, por ejemplo). Gráfica sola indica que la gráfica no aparece junto a otra para comparar, y Sin explicación indica que la explicación de la gráfica se encuentra en otra página.

### Datos y diseño

Se seleccionó una muestra aleatoria simple de tesis de Actuaría registradas en la biblioteca con fecha entre 1990 y 2007 (Abril 2008). No se pudieron encontrar 3 de las 125 tesis seleccionadas en la muestra. Se registraron todas las gráficas (separadas en gráficas de datos y gráficas analíticas según criterios establecidos) contenidas en cada tesis seleccionadas, y cada una se clasificó según su tipo y su contenido de basura gráfica. Los criterios para basura gráfica son apreciativos: los estudiantes que levantaron los datos fueron primero entrenados para regularizar en lo más posible los criterios.