



## Trabajo Práctico N° 1

### Visualizacion

Presentado en la fecha: 26/08/2023

Hecho por:

Huarca Brian

Nicolas Benitez

Facundo Rodriguez

Derlis Walter Hodge

Yasmin Salvi

# Contents

<b>Resumen</b>	<b>2</b>
<b>Objetivo</b>	<b>3</b>
<b>Desarrollo</b>	<b>4</b>
<b>1 Enseñanzas de Alberto Cairo</b>	<b>4</b>
1.1 La importancia de la funcionalidad en una Visualizacion . . . . .	4
1.2 No hay que entender la Infografía como arte . . . . .	6
1.3 La tecnologia y el no saber dibujar no debe ser un obstaculo . . . . .	8
<b>2 Visualizacion Engañosa</b>	<b>9</b>
2.1 Caso de Mortalidad - Provincia de BS.AS . . . . .	9
2.2 Caso de Cemento Portland - Anuario 2017 . . . . .	11
<b>3 Casos de Buena Visualizacion</b>	<b>13</b>
3.1 Estimaciones Agricolas de la Provincia de Buenos Aires . . . . .	13
3.2 Actividad Productiva de determinados sectores industriales . . . . .	14
<b>4 Visualizacion en R</b>	<b>15</b>
<b>Conclusión</b>	<b>16</b>
<b>Anexo</b>	<b>17</b>

# Resumen

El presente informe busca introducir al lector a las buenas practicas y falencias que pueden encontrarse en el analisis de visualización. Tambien sobre las características fundamentales que debe tener.

Como detalle final se modela un grafico en R sobre los casos por tipo de defunción en el año 2011 del municipio de Merlo.

# Objetivo

En el presente trabajo se propone una introducción para entender y comprender los aspectos positivos que debe tener una infografía, visualización, o gráfico, a partir de ejemplos concretos de buenos y malos diseños, tomando en consideración las enseñanzas que nos deja Alberto Cairo.

# Enseñanzas de Alberto Cairo

## 1.1 La importancia de la funcionalidad en una Visualizacion

El autor plantea que el diseño debe estar dirigido por las funciones que ese grafico debe tener. El Arte Funcional.

El siguiente ejemplo es un claro caso de una buena idea, pero mala funcionalidad:

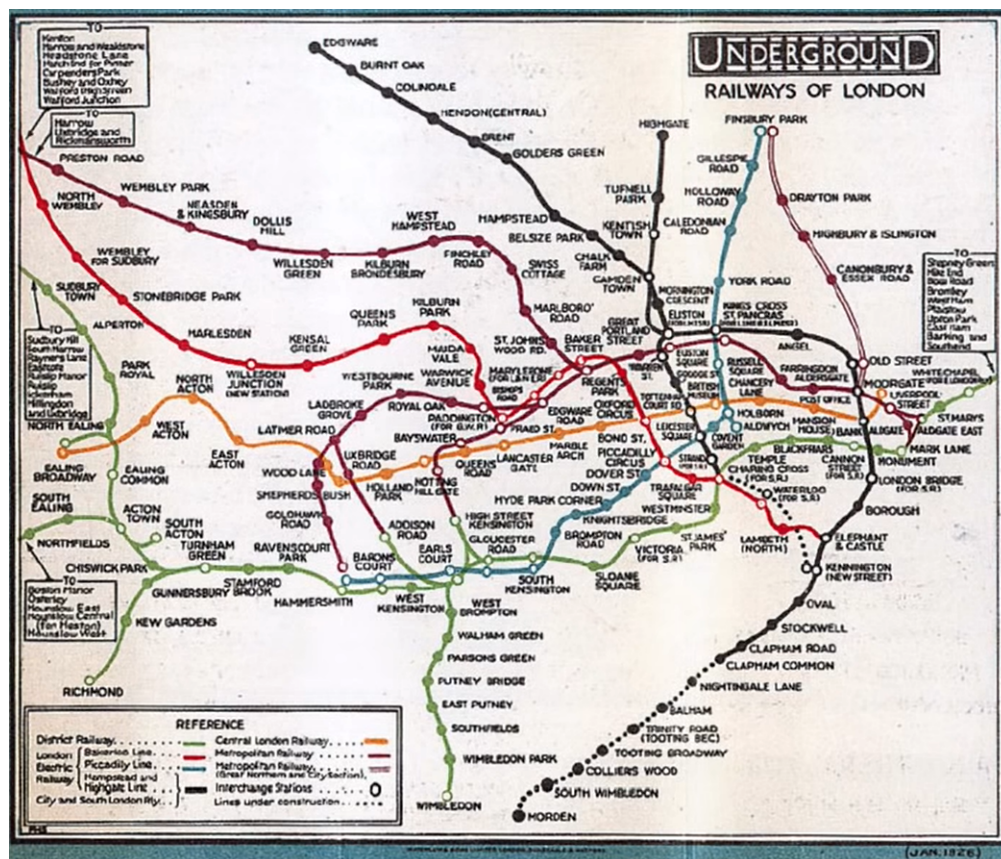


Figure 1: Visualizacion de premetro

El siguiente caso corrige el caso anterior agregando la funcionalidad faltante:



Figure 2: Visualizacion de premetro 2

## 1.2 No hay que entender la Infografía como arte

El autor plantea en su libro que no hay que entender la infografía como arte ya que estos 2 tienen distintos enfoques distintos, el arte es algo subjetivo donde el artista busca expresar sus sentimientos y ser bello.

Un ejemplo de una bella obra es "La Virgen de la Roca" del artista Leonardo Da Vinci.



Figure 3: Visualización de La Virgen de la Roca



Mientras que la Infografía tiene un enfoque objetivo, estadístico y específico donde la misma debe basarse en observaciones reales y ser fiel a esas observaciones sin el objetivo de ser bello aunque esto no quita no pueda serlo.

Ejemplo de ello puede ser una de las paginas del libro de notas de Da Vinci ya que el ademas de ser un artista tambien era un ingeniero y realizaba investigaciones.

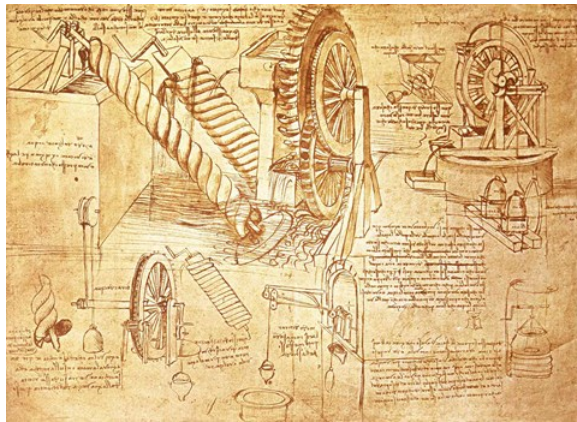


Figure 4: Visualizacion de Libro de Notas de Da Vinci



### 1.3 La tecnología y el no saber dibujar no debe ser un obstaculo

El autor plantea que un principal obstaculo en el periodismo de datos es la idea de que se necesitan conocimientos en software, programacion y saber dibujar para poder llevar a cabo la infografia, pero el nos demuestra lo contrario con Visualizaciones de antaño que fueron creadas mucho antes de que estas tecnologias existieran y sin tener una gran habilidad en las artes.

Esta Visualizacion fue realizado en un mapa del distrito Soho en Londres por el Medico John Snow en 1854, el cual cada linea negra representa una residencia con un fallecido por colera, el Medico noto que habia un patron en donde se concentraba la gran mayoria de fallecidos por "Colera" en el centro de la ciudad donde en su investigacion llego a la conclusion de que el agua del pozo central estaba contaminada.



Figure 5: Visualizacion de Mapa de Colera de John Snow

# Visualizacion Engañososa

## 2.1 Caso de Mortalidad - Provincia de BS.AS

El siguiente gráfico de columnas apiladas muestra la evolución de la mortalidad en la provincia de Buenos Aires y sus causas más frecuentes. Fuente: CLICK AQUI.

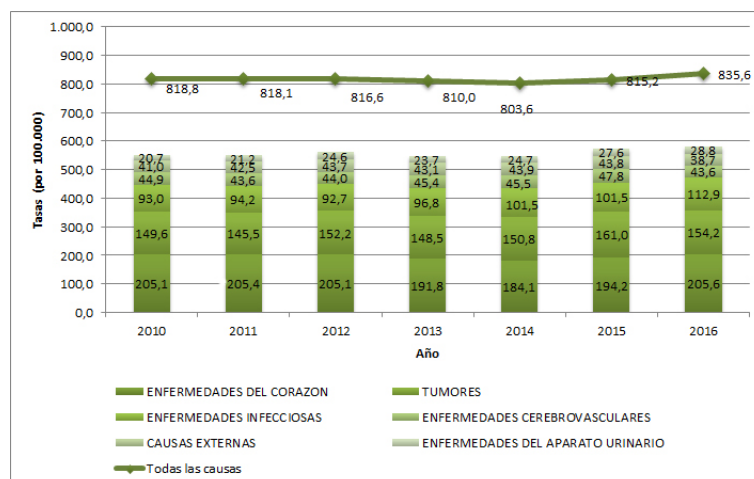


Figure 6: Mortalidad por causas mas frecuentes

Aspectos erroneos:

1. El total de los casos por Mortalidad no se distingue por cada barra (año), y ademas el indice del eje Y confunde las cantidades acumuladas.
2. Es dificil crear una comparacion entre las distintas enfermedades
3. En cada columna se representan las diferentes causas de mortalidad en distintas gamas del mismo color, esto dificulta identificar cada sección. Utilizar diferentes colores facilitaria comparar las variables.

4. El significado de la línea con marcadores no es del todo claro. Si bien sirve para mostrar una tendencia de los datos, los valores de los marcadores no coinciden con el total de cada columna.

Aspectos a mejorar:

1. Indicar el total de cada columna
2. Separar cada mortalidad por barra
3. Detallar colores distintivos por cada mortalidad

## 2.2 Caso de Cemento Portland - Anuario 2017

El siguiente grafico corresponde a uno de los anuarios que realiza la AFCP (Asociacion de Fabricantes de cemento Portland), es una entidad civil sin fines de lucro, fundada en 1922, que representa a la industria argentina del cemento portland.

El "cemento Portland" es un tipo de cemento hidráulico que se utiliza ampliamente en la construcción debido a sus propiedades y versatilidad. Es uno de los tipos de cemento más comunes y estándares que se utilizan en todo el mundo para la producción de hormigón. La industria del cemento en Argentina está compuesta por cuatro productores: • Loma Negra C.I.A.S.A • Holcim (Argentina) S.A. • Cementos Avellaneda S.A. • PCR S.A. Fuente: [CLICK AQUI](#).

Aspectos erroneos:

1. Hay etiquetas (Clinter y Blanco) que estan mencionadas pero que no se representan en el grafico de torta. Hay que eliminar la redundancia
2. No hay gran diferencia en las tonalidades de cada componente de cemento.
3. No se visualizan etiquetas en cada porcion del grafico para mostrar el porcentaje correspondiente
4. Se mezcla Clinker siendo un componente para la fabricación de cemento con tipos de cemento (blanco y otros)

### EXPORTACIONES 2017

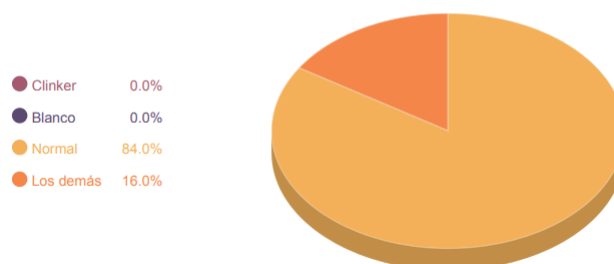


Figure 7: Exportaciones

Aspectos a mejorar:

1. Eliminar la redundancia de los componentes y tipos de cemento que no se usan en el grafico
2. Mejorar las tonalidades en los colores
3. Detallar mediante etiquetas a que corresponde cada componente y/o cemento.

# Casos de Buena Visualizacion

## 3.1 Estimaciones Agrícolas de la Provincia de Buenos Aires

En esta Visualización interactiva podemos encontrar datos referentes a los cultivos principales de la Provincia de Buenos Aires, se tienen totales de producción por año y tipo de cultivo. También nos permite comparar la producción, y obtener una idea de tendencia por tipo de cultivo. Fuente: [CLICK AQUI](#).

### Estimaciones Agrícolas de la Provincia de Buenos Aires

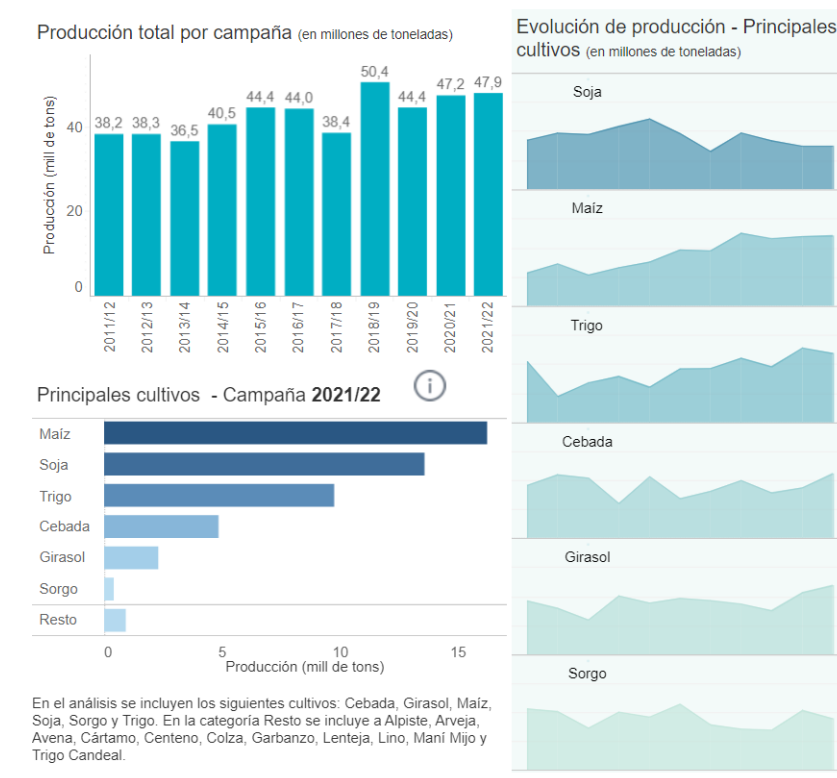


Figure 8: Grafico de Estimaciones Agrícolas

## 3.2 Actividad Productiva de determinados sectores industriales

El siguiente grafico nos permite entender el nivel de productividad por sector industrial en la Provincia de Buenos Aires. El mismo detalla por dimension de tiempo anual, trimestral y mensual. A su vez detalla en cada periodo las cantidades totales producida por sector. Finalmente si apoyamos el cursor sobre cada barra obtenemos el detalle de variacion acumulada con respecto al año anterior. Fuente: [CLICK AQUI](#).

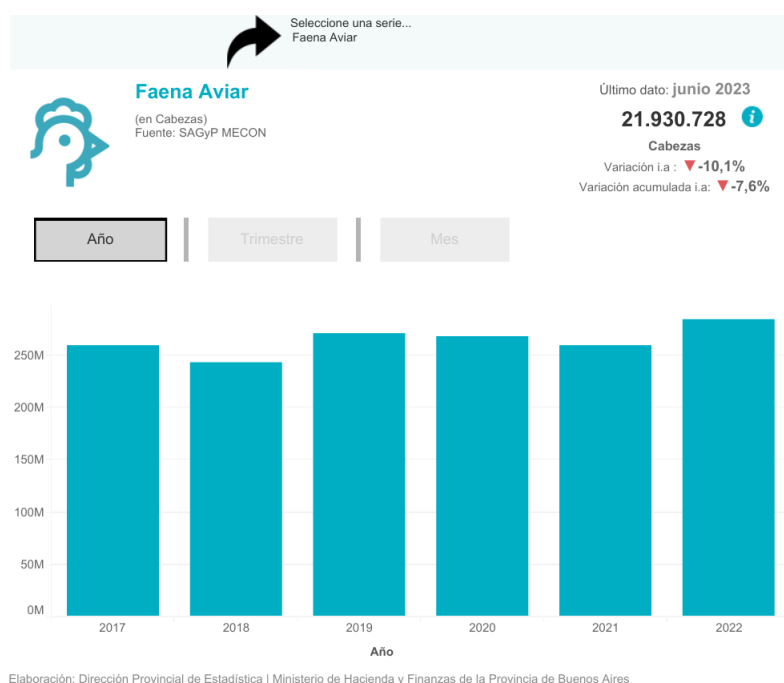


Figure 9: Pcia. Bs.As - Actividad Productiva por Faena Aviar



## Visualizacion en R

A partir del siguiente proyecto en R, y mediante un grafico de barras podemos tener un panorama de como venian los casos por defunciones registradas en el municipio de Merlo a 2019.

## Conclusión

El presente trabajo nos permitio entender la importancia de una visualización, a partir de las manipulacion, falencias, criterios, como tambien apoyanos en herramientas de la actualidad que nos facilitan recorrer un detalle en los datos para sacar conclusiones mas exactas para poder tomar decisiones. Por otro lado, podemos mencionar algunas enseñanzas a la hora de plasmar una visualización tales como: El objetivo que debe perseguir; relativizar los datos en función de lo que se mida; que sean amigables para que alguien no es especialista pueda entenderlo;

```
install.packages("readr")
install.packages("sqldf")
install.packages("dplyr")
install.packages("plotly")

library(readr)
library(sqldf)
library(dplyr)
library(plotly)

#
-----

#01.IMPORTACION DE DATOS
rm(defunciones)
defunciones <- read.table(file.choose(),
                          sep=',',
                          header=TRUE,
                          fill = TRUE,
                          row.names = NULL,
                          stringsAsFactors = FALSE)

View(defunciones)
str(defunciones)
names(defunciones)
head(defunciones)

#
-----

#02. CAMBIO DE NOMBRE EN COLUMNAS
defunciones <- defunciones %>% rename (municipio_nom = municipio_nombre .,
                                       causa_muerte_agrup = causa_muerte_capitulo,
                                       causa_muerte_detalle = causa_muerte_agrupamiento)

#
-----

#03. REDUCCION DEL DATASET
rm(defunciones_red)
defunciones_red = subset(defunciones, select = c(anio, municipio_nom,
                                                causa_muerte_agrup, defunciones_cantidad
))
```

```

36 View(defunciones_red)
37 str(defunciones_red)
38 defunciones_red %>% distinct(anio)
39 #
-----

40 #04. CASTEO LOS DATOS
41 #defunciones_red$sexo <- as.factor(defunciones_red$sexo)
42 #defunciones_red$grupo_edad <- as.factor(defunciones_red$grupo_edad)
43 #defunciones_red$causa_muerte_detalle <- as.factor(defunciones_red$causa_muerte_detalle)
44 defunciones_red$anio <- as.factor(defunciones_red$anio)
45 defunciones_red$municipio_nom <- as.factor(defunciones_red$municipio_nom)
46 defunciones_red$causa_muerte_agrup <- as.factor(defunciones_red$causa_muerte_agrup)
47 defunciones_red$defunciones_cantidad <- suppressWarnings(as.numeric(defunciones_red$
  defunciones_cantidad))
48 #
-----

49 #05. COMO BUSCO LOS VALORES RAROS PARA LUEGO QUITARLOS?:
50 #pruebal %>% distinct(causa_muerte_detalle)
51
52 #EXPLORACION DE DATOS: CUANTOS DATOS CON VALOR "IGNORADO TENGO"?
53 sqldf('
54 SELECT anio, municipio_nom, count(*)
55 FROM pruebal
56 WHERE municipio_nom = "NA"
57 GROUP BY anio, municipio_nom
58 ORDER BY 3 DESC
59 ')
60
61 #CONVIERTO LOS VALORES "IGNORADOS" (MUNICIPIO NOMBRE) y "SIN DATO" (EDAD) A NA PARA
  ELIMINARLOS
62 #rm(pruebal)
63 #pruebal <- defunciones_red
64 defunciones_red[defunciones_red == "Ignorado"] <- NA
65 defunciones_red[defunciones_red == "Sin dato"] <- NA
66 defunciones_red[defunciones_red == ""] <- NA
67 View(defunciones_red)
68
69 # CUANTOS NA TENGO POR COLUMNA?
70 View(summarise_all(defunciones_red, funs(sum(is.na(.)))))
71
72 #ELIMINO LOS NA
73 defunciones_red <- na.omit(defunciones_red)
74
75 # CUANTOS NA TENGO POR COLUMNA? (REPASO)
76 View(summarise_all(defunciones_red, funs(sum(is.na(.)))))
77 #
-----

78 #FILTRO POR ANIO
79 #prueba <- defunciones_red %>% filter(between(anio, 2010, 2019))
80 #prueba %>% distinct(anio)
81 #prueba %>% group_by(anio, municipio_nombre) %>% summarise(cantidad = n()) %>% arrange
  (desc(cantidad))

```

```
82 #
-----

83 #05. FILTRO POR MERLO Y ANIO 2019
84 defunciones_red %>% distinct(municipio_nom)
85 names(defunciones_red)
86
87 rm(defunciones_red_merlo)
88 defunciones_red_merlo <- filter(defunciones_red, municipio_nom == "Merlo" & anio == 2019)
89   %>%
90   select(causa_muerte_agrup, defunciones_cantidad)
91 defunciones_red_merlo <- filter(defunciones_red_merlo, defunciones_cantidad > 0, Mac2 >
92   0, Mac3 > 0, Mac4 > 0)
93 summary(defunciones_red_merlo)
94
95 #AGRUPAMOS:
96 defunciones_red_merlo <- defunciones_red_merlo %>% group_by(causa_muerte_agrup) %>%
97   summarise(across(c(defunciones_cantidad), sum))
98
99 #ENF. DEL SISTEMA OSTEOMUSCULAR Y DEL TEJIDO CONJUNTIVO
100 #EMBARAZO, PARTO Y PUERPERIO
101
102 defunciones_red_merlo = filter(defunciones_red_merlo, defunciones_cantidad>10)
103 rm(defunciones)
104 rm(defunciones_red)
105
106 View(summarise_all(defunciones_red_merlo, funs(sum(is.na(.)))))
107 str(defunciones_red_merlo)
108
109 write.csv(defunciones_red_merlo, "datos.csv")
110
111 datos <- read.table(file.choose(),
112   sep=',',
113   header=TRUE,
114   fill = TRUE,
115   row.names = NULL,
116   stringsAsFactors = FALSE)
117 #
-----

118 #06 VISUALIZACION
119
120 library(plotly)
121
122 fig <- plot_ly(datos,
123   x = datos$causa_muerte_agrup,
124   y = datos$defunciones_cantidad,
125   name = "Causas de Muertes en Pcia. de Buenos Aires 2019",
126   type = "bar"
127 )
128
129 fig
130
```

```
131 plot(datos)
```