

# INSTITUTO TECNOLÓGICO DE OAXACA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

## **TAREA**

MINERIA DE DATOS RSTUDIO

## **MATERIA:**

TOPICOS DE CIENCIAS DE LOS DATOS.

## **DOCENTE:**

HERNÁNDEZ ABREGO ANAYANSI CRISTINA

## **INTEGRANTE:**

ARAGÓN RAMÍREZ ANGEL DANIEL

**GRUPO:** ISU    **HORA:** 7:00 – 8:00 HRS.

Se obtendrá los 10 nombres más usados de los años no importa el sexo.

Usamos para que nos diga los nombres.

```
arrange(babyNa, desc(Count))
babyNa %>% filter(Name=="Christopher")
```

Posteriormente filtramos los 10 nombres más usados en los años.

```
arrange(babyNa, desc(Count))
8 babyNa %>% filter(Name=="Christopher")
9 babyNa %>% filter(Name=="Daniel")
10 babyNa %>% filter(Name=="David")
11 babyNa %>% filter(Name=="James")
12 babyNa %>% filter(Name=="Jennifer")
13 babyNa %>% filter(Name=="John")
14 babyNa %>% filter(Name=="Matthew")
15 babyNa %>% filter(Name=="Michael")
16 babyNa %>% filter(Name=="Robert")
17 babyNa %>% filter(Name=="William")
```

Hacemos otra filtración para comparar los nombres y nos muestre en una sola gráfica.

```
nombre <- filter(babyNa, Name=="Christopher" | Name=="Daniel" | Name=="David" | Name=="James" | Name=="Jennifer" | Name=="John" |
  Name=="Matthew" | Name=="Michael" | Name=="Robert" | Name=="William")
```

Graficamos con un ggplot la variable nombre, la cual tendrá la forma de la gráfica en nuestro caso usamos geom\_point la cual tendrá un color steelblue y un tamaño de 1, junto que estará con otro geom\_point la cual tendrá el mismo color.

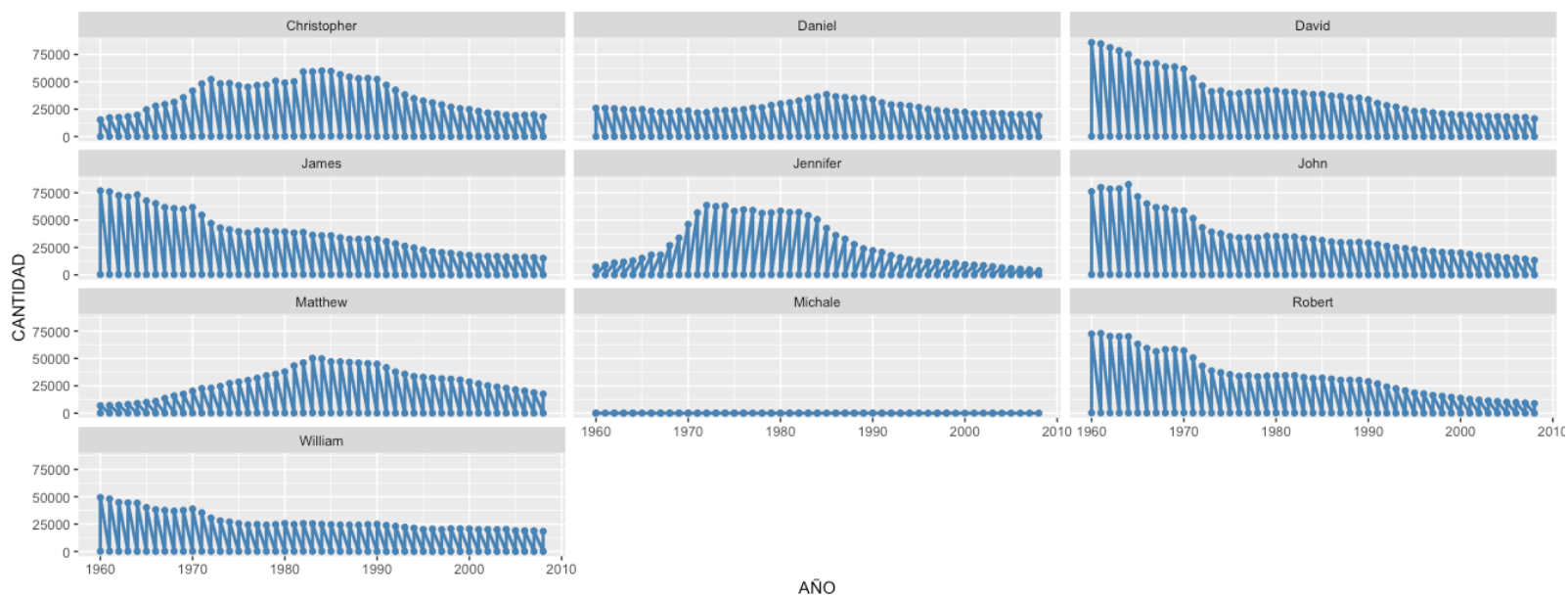
```
ggplot(nombre, aes(x=Year, y=Count))+ geom_line(color = "steelblue", size = 1) +
  geom_point(color = "steelblue") + facet_wrap(~ Name, ncol = 3)+
  labs(title = "Los 10 nombres más populares en 1960-2007")
```

Posterior mente se incorporara el facet\_wrap, la cual tendra un labs la cual tendra un titulo y en las coordenada la descripcion.

```
labs(title = "Los 10 nombres más populares en 1960-2007",  
      subtitle = "",  
      caption = "",  
      x = "AÑO", y = "CANTIDAD",  
      fill = NULL)
```

## Grafica

Los 10 nombres más populares en 1960-2007



## Conclusiones

En este trabajo pusimos en practica lo aprendido en la clase ya que hay muchas formas en como hacer la practica se puede usar sacando el rank de los datos asi como las filtraciones de las fechas de igual manera por vectores. Ya que en Rstudio nos ayuda para modifica nuestros datos y poder sacar estadisticas en forma de graficas y asi visualir los datos.

En esta practica nos ayuda a no olvidar el uso de Rstudio ya que para mi es muy interancen el uso de R.