



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

Actividad Práctica I

Rubén Soza
Computación Estadística con R

22 de Agosto de 2019

Instrucciones

Instale los siguientes paquetes: `babynames`, `nycflights13`, `maps` y carguelos (con `library`) junto a `tidyverse`. Resuelva las siguientes preguntas:

Pregunta 1

Considere la tabla `births` paquete `babynames`.

- (a) ¿Cuántas columnas y filas posee esta tabla?
- (b) Utilizando un gráfico adecuado, estudie los nacimientos por años en EE.UU. ¿En qué década se obtuvo la menor cantidad de nacimientos?

```
library(babynames)
library(nycflights13)
library(maps)
```

```
data("births")
```

```
# Pregunta 1 #
glimpse(births)
```

```
## Observations: 109
## Variables: 2
## $ year    <int> 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1...
## $ births  <int> 2718000, 2777000, 2809000, 2840000, 2869000, 2966000, 2...
```

```
# Pregunta 2 #
```

```
ggplot(data = births) +
  geom_line(aes(x = year, y = births, col = "darkblue")) + theme(plot.subtitle = element
  plot.caption = element_text(vjust = 1),
  axis.line = element_line(linetype = "solid"),
```

```

panel.grid.major = element_line(linetype = "blank"),
axis.text = element_text(size = 12),
axis.text.x = element_text(size = 12),
panel.background = element_rect(fill = NA,
  linetype = "solid"), legend.key = element_rect(fill = "white"),
legend.background = element_rect(fill = NA,
  linetype = "solid"))

```



```

births %>%
  filter(year >= 1910) %>%
  mutate(decade = floor(year/10)*10) %>%
  group_by(decade) %>%
  summarise(nacimientos = sum(births)) %>%
  filter(nacimientos == min(nacimientos))

```

```

## # A tibble: 1 x 2
##   decade nacimientos
##   <dbl>      <int>
## 1  1930      24374000

```

Pregunta 2

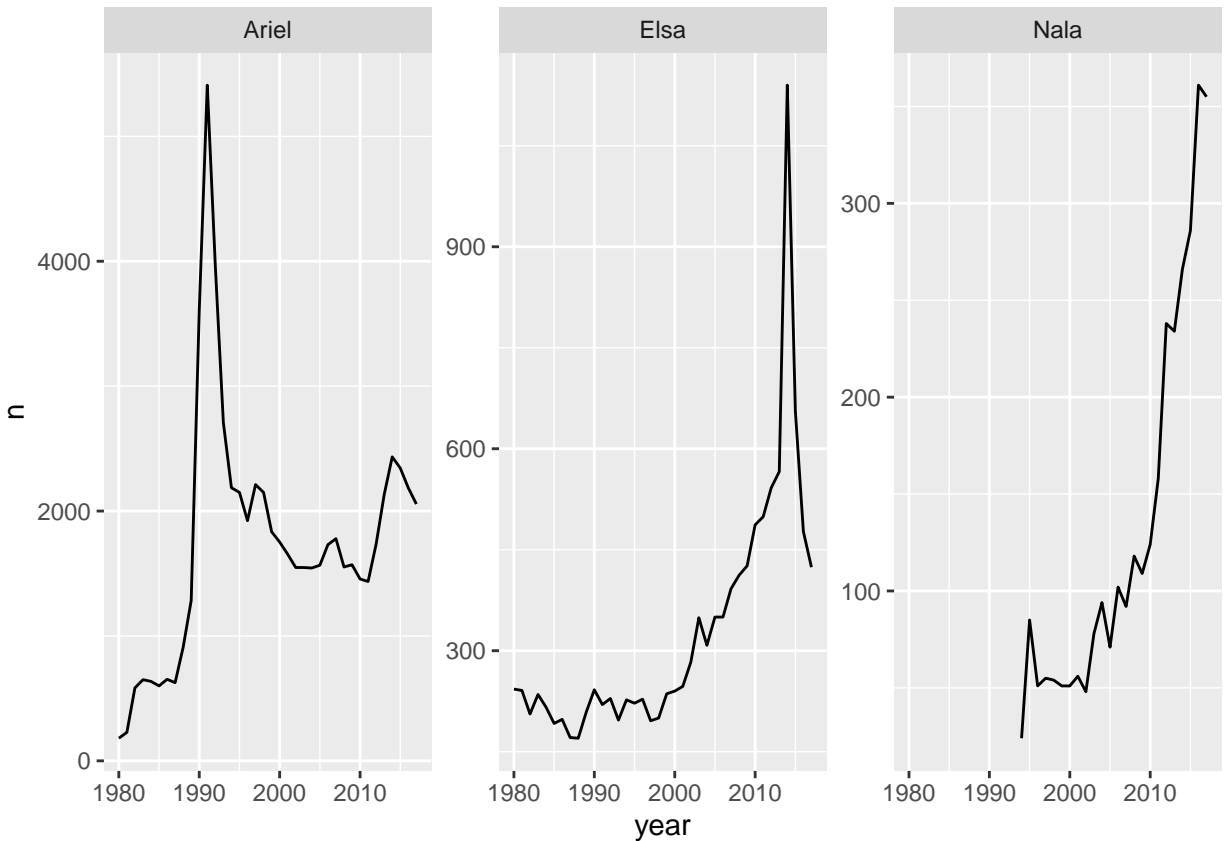
Considere ahora la tabla `babynames` del paquete `babynames`.

- (a) Realice el proceso necesario para obtener la misma estructura en cuanto a **columnas** que la tabla `births`, esto es, generar una nueva tabla que nos entregue el número de nacimientos por año.
- (b) Genere un data frame partiendo de la tabla `babynames` que contenga los nacimientos de las personas de género femenino con el nombre **Nala**, **Ariel** y **Elsa** desde los años 1980.
- (c) Con el data frame obtenido en la parte anterior, genere un gráfico que contenga la información de los nacimientos por año de cada uno de los nombres mencionados y mencione una hipótesis/suposición al respecto de lo observado. *Hint*: Use `face_wrap(~ name, scales = "free_y")`.

```
# Parte a #
data("babynames")
babynames %>%
  group_by(year) %>%
  summarise(births = n() ) -> births2

# Parte b #
babynames %>%
  filter(sex == "F", year >= 1980, name %in% c("Nala","Ariel","Elsa")) %>%
  arrange(name) -> births3

# Parte c #
ggplot(data = births3) +
  geom_line(aes(x = year, y = n)) +
  facet_wrap( ~ name, ncol = 3, nrow = 1, scales = "free_y")
```



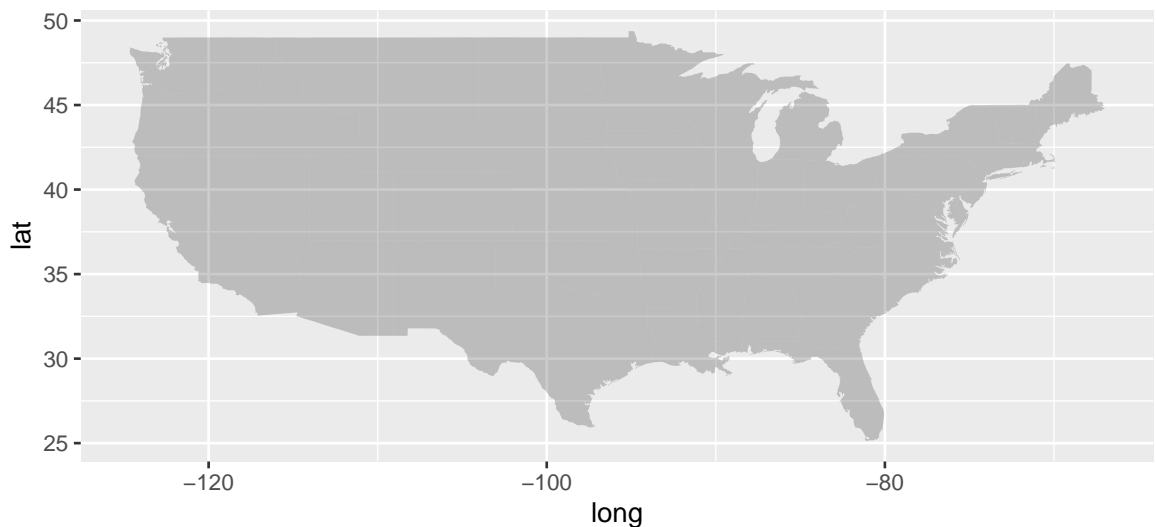
Pregunta 3

- Utilizando las tablas `airports` y `flights` del paquete `nycflights13` obtenga una tabla que contenga conteos de vuelos según su destino `dest` además de la longitud y latitud del aeropuerto(de destino).
- Apoyándose del siguiente gráfico:

```
library(maps)
us <- map_data("state")

ggmap <- ggplot() +
  geom_polygon(data = us, aes(long, lat, group = group), alpha = 0.25) +
  coord_fixed() # esto es para mantener la razón 1:1

ggmap
```



1. Agregue una capa de puntos ubicando los aeropuertos obtenidos de la pregunta anterior y usando además: `aes(size = la_cantidad_de_vuelos_a_dicho_aeropuerto)`.
2. A la izquierda del gráfico anterior se observan 2 puntos. Genere el/los pasos necesarios para seleccionarlos usando la tabla resultante de la pregunta 5 para identificar los nombres de dichos aeropuertos y responda: ¿Donde están? ¿Qué gracias tienen?

```
library(nycflights13)
data("airports", package = 'nycflights13')
data("flights", package = 'nycflights13')
glimpse(airports)
```

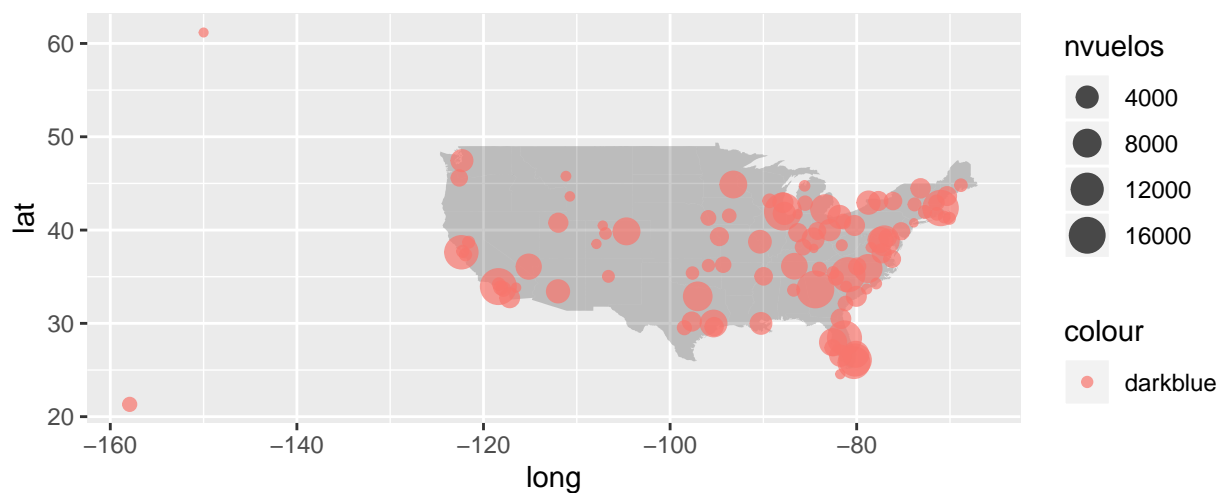
```
## Observations: 1,458
## Variables: 8
## $ faa   <chr> "04G", "06A", "06C", "06N", "09J", "0A9", "0G6", "0G7", ...
## $ name  <chr> "Lansdowne Airport", "Moton Field Municipal Airport", "S...
## $ lat   <dbl> 41.13047, 32.46057, 41.98934, 41.43191, 31.07447, 36.371...
## $ lon   <dbl> -80.61958, -85.68003, -88.10124, -74.39156, -81.42778, -...
## $ alt   <int> 1044, 264, 801, 523, 11, 1593, 730, 492, 1000, 108, 409,...
## $ tz    <dbl> -5, -6, -6, -5, -5, -5, -5, -5, -5, -8, -5, -6, -5, -5, ...
## $ dst   <chr> "A", "A", "A", "A", "A", "A", "A", "A", "U", "A", "A", "...
## $ tzone <chr> "America/New_York", "America/Chicago", "America/Chicago"...
```

```
# a)
flights %>%
  left_join(airports, by = c("dest" = "faa")) %>%
  group_by(name) %>%
  summarise(nvuelos = n()) %>%
  left_join(airports, by = "name") %>%
  select(name,lat,lon,nvuelos) -> airports2

# b)
ggmap <- ggplot() +
  geom_polygon(data = us, aes(long, lat, group = group), alpha = 0.25) +
  coord_fixed()

ggmap + geom_point(data = airports2, aes(x = lon, y = lat, size = nvuelos, color = "darkblue"))

## Warning: Removed 1 rows containing missing values (geom_point).
```



```
# c)
airports2 %>%
  filter(lon < -130) # Hawaii y Alaska :o

## # A tibble: 2 x 4
```

##	name	lat	lon	nvuelos
##	<chr>	<dbl>	<dbl>	<int>
## 1	Honolulu Intl	21.3	-158.	707
## 2	Ted Stevens Anchorage Intl	61.2	-150.	8