

# ITAM - Estadística 1

## Sheet 1 — Assignment 01

### Básicos Exploratory Data Analysis

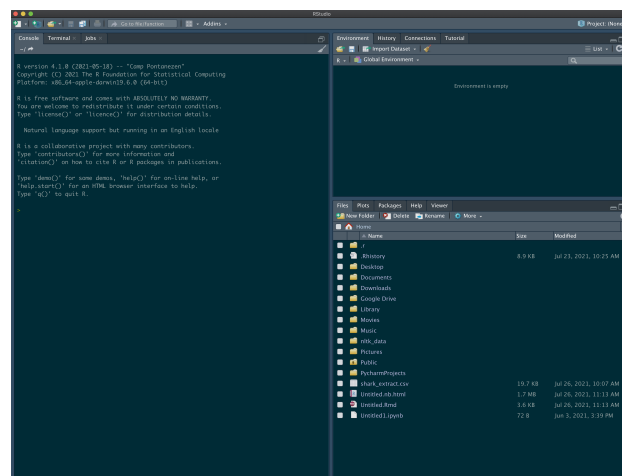
Este laboratorio nos servirá para tener un primer acercamiento a R y a sus principales funciones para realizar **Exploratory Data Analysis (EDA)**. El Exploratory Data Analysis (EDA) es utilizado constantemente como primer paso en la exploración de datos, ya sean cualitativos o cuantitativos para cualquier escala de medición. En la tarea 1 se explora el dataset **Sharks** obtenido a través de Kaggle <https://www.kaggle.com/felipeesc/shark-attack-dataset> y de autoría de Shark Research Institute. Sharks es un dataset que contiene un subconjunto de datos de ataques de tiburones en todo el mundo. La recopilación de estos datos se realizó con la finalidad de encontrar comportamientos de las especies y encontrar patrones que llevarán al ataque de tiburones a personas.

Solamente dos docenas de especies de tiburones son considerados letales para los humanos por su tamaño y dentadura. Es de notar que cada registro marcado como ataque no necesariamente implican un daño a un humano, ya que si solamente se muerde alguna herramienta (por ejemplo un surfboard) se considera ataque.

1. **Instalación de R y RStudio** Para poder correr en notebook adjunto, se debe instalar R y Rstudio.

- Para instalar **R** ir a la página: <https://cran.itam.mx/> Se selecciona el sistema operativo que tenga en su computadora y se siguen los pasos de instalación
- Para instalar **RStudio** ir a la página: <https://www.rstudio.com/products/rstudio/> y descargar la versión open source.

Una vez instalados ambos debe aparecer una página como la siguiente (puede ser blanco en lugar de azul):



2. **Instalar paquetes necesarios** R es un software que requiere instalar paquetes. El paquete que vamos a requerir para este primer laboratorio es tidyverse:

- Instalar tidyverse: En la consola poner **install.packages('tidyverse')**
- Cargar tidyverse: En la consola poner **library(tidyverse)**. Este paso debe repetirse cada vez que se inicie R.

3. **Correr el laboratorio**

- Para correr laboratorio, abre en Rstudio el archivo **Assignment01.Rmd**. Asegurate de contar con el archivo **shark\_extract.csv** en la misma ruta.
- El notebook tiene chunks que puedes ir corriendo uno a uno. Asegurate de correr toda la primer sección
- Basado en los resultados que te dieron, contesta el quiz en Canvas