



# R para Economía y Negocios

## TAREA 1

Profesor: *Victor Macías E.*  
Ayudante: *Gabriel Cabrera G.*<sup>1</sup>  
27 agosto 2018

---

### Instrucciones

1. Esta tarea debe ser entregada en grupos de máximo 3 personas.
2. Se debe enviar una carpeta comprimida que contenga: un archivo .Rproj (proyecto en R), un script y un breve informe en R Markdown con sus resultados, incluyendo tablas, gráficos, etc., según corresponda. El formato del informe puede ser word o pdf.
3. El asunto del email con su tarea debe ser “Tarea 1-R 1-Apellido 1-Apellido 2-Apellido 3”
4. La fecha de entrega es el Jueves 13 de Septiembre del 2018 hasta las 23:59 hrs al correo del curso: [r2018uchile@gmail.com](mailto:r2018uchile@gmail.com).

---

“Data is a precious thing and will last longer than the systems themselves”.

*Tim Berners-Lee*, inventor del the World Wide Web

---

Una importante consultora de marketing está en la búsqueda de estudiantes con conocimiento del lenguaje R para realizar un estudio respecto a la industria del cine. Usted piensa que es una muy buena oportunidad postular, considerando que está cursando un curso en el que está aprendiendo R. Después de haber pasado varias pruebas y entrevistas, usted es contratado y como primera tarea le entregan la carpeta *Datasets* que contiene:

Table 1: Breve descripción de las bases de datos.

Base de datos	Descripción
<i>genres.csv</i>	Géneros de cada película.
<i>keywords.csv</i>	Palabras claves de cada película.
<i>production_companies.csv</i>	Compañía a cargo de la producción.
<i>production_countries.csv</i>	Países donde se realizó la producción.
<i>movie_dataset.csv</i>	Base de datos principal.

A continuación le piden los siguientes **data frames** (sin considerar los *missing values*) pensando que en un futuro no tan distante la información que contendrán será presentada gráficamente:

---

<sup>1</sup> ✉: [gcabrerag@fen.uchile.cl](mailto:gcabrerag@fen.uchile.cl)

1. Identificar las 10 palabras claves (*keywords*) que más se repiten. (10 puntos)
2. Identificar:
  - a. Las 10 películas con los ingresos (*revenue*) más altos. (10 puntos)
  - b. Las 10 películas con los ingresos (*revenue*) más bajos. (10 puntos)
  - c. El número de películas con ingresos igual a cero. (10 puntos)
3. Identificar la cantidad de películas por país, luego calcular la utilidad<sup>2</sup> (*revenue - budget*) y encontrar las 5 países con mayor y menor utilidad promedio. (10 puntos)
4. Dado que últimamente las salas IMAX han tenido una muy buena recepción por parte del público, la consultora le pide calcular el ingreso (*revenue*) promedio por género, pero sólo de aquellas películas que fueron estrenadas en formato IMAX. (15 puntos)
5. La industria del cine ha cambiado mucho desde sus inicios y miles de películas son estrenadas cada año. Considerando sólo las películas que se encuentran en las bases de datos, identifique el porcentaje de películas estrenadas por década<sup>3</sup>. (15 puntos)
6. Identifique la cantidad de películas por estado de Estados Unidos (vea el archivo *estados.csv*) y el presupuesto promedio (*budget*) por estado. ¿Qué estados nunca fueron usados como lugar de filmación?, ¿Existe algún estado con presupuesto cero?. (15 puntos)

---

<sup>2</sup>Asuma que el presupuesto fue el gasto total.

<sup>3</sup>Una librería útil para trabajar con fechas es `lubridate`. La década se puede construir con la ayuda del operador `%%`.