Evaluación

## Consideraciones

* Para esta evaluación usted debera entregar un script de R con los comandos necesarios para responder cada uno de los puntos.
* Además se evaluará la calidad (no necesariamente la cantidad) de comentarios en las respuestas.
* Esta evaluación se puede hacer en grupo de máximo 3 personas.
* La forma de entrega será vía email a [**jbkunst@gmail.com**](mailto:jbkunst@gmail.com) con plazo el día viernes 7 de diciembre.

## Ejericicio 1

Considere los siguientes datos de alumnos encuestados al final de semestre sobre un curso de R:

library(tidyverse)  
  
alumnos <- read\_csv("https://goo.gl/Wy3GsU")  
glimpse(alumnos)

## Observations: 3,000  
## Variables: 3  
## $ dificultad <dbl> 39.64, 50.66, 36.79, 69.86, 52.64, 37.00, 54.79, ...  
## $ satisfaccion <dbl> 62.78, 57.49, 67.23, 44.47, 43.84, 64.77, 65.74, ...  
## $ grupo <chr> "grp04", "grp03", "grp04", "grp02", "grp02", "grp...

A cada alumno encuestado se le preguntó sobre su satisfacción y percepción de la duración del curso.

* Obtenga un diagrama de puntos entre las variables duracion y satisfaccion.
* Comente respecto al resultado. Intente realizar una hipótesis respecto a lo que observa.
* Al gráfico anterior agregue una línea de tendencia usando la funcion geom\_smooth (Puede revisar ejemplos, ejecutar ?geom\_smooth)

Ahora suponga que el equipo académico le recomienda considerar el grupo de donde el alumno proviene para así seguir estudiando la evidencia encontrada en el punto anterior:

* Vuelva a hacer un diagrama de puntos pero ahora diferenciando mediante el color al grupo al que el alumno pertenece.
* Al grafico anterior agregue una línea de tendendencia y además utilice el comando facet\_wrap para crear graficos de acuerdo al grupo del alumno.
* Con lo obtenido, ¿Qué sucede con la hipótesis obtenida anteriormente?

## Ejercicio 2

En esta parte trabajaremos con datos de variaciones de temperatura durante los años:

temperatura <- read\_csv("https://goo.gl/3JUCma")  
temperatura

## # A tibble: 1,992 x 5  
## anio\_mes mediana inferio superio decada  
## <date> <dbl> <dbl> <dbl> <int>  
## 1 1850-01-01 -0.702 -1.10 -0.299 1850  
## 2 1850-02-01 -0.281 -0.673 0.117 1850  
## 3 1850-03-01 -0.732 -1.08 -0.382 1850  
## 4 1850-04-01 -0.569 -0.904 -0.237 1850  
## 5 1850-05-01 -0.326 -0.662 0.005 1850  
## 6 1850-06-01 -0.212 -0.515 0.085 1850  
## 7 1850-07-01 -0.128 -0.457 0.199 1850  
## 8 1850-08-01 -0.232 -0.595 0.132 1850  
## 9 1850-09-01 -0.435 -0.808 -0.063 1850  
## 10 1850-10-01 -0.452 -0.793 -0.104 1850  
## # ... with 1,982 more rows

* Relice un gráfico de línea considerando las columnas anio\_mes y mediana
* Ahora cargue el paquete lubridate, y usando las funciones year y month sobre la columna anio\_mes cree dos variables en la tabla temperatura y nómbrelas como anio y mes.
* Revise y considere los ejemplo de geom\_raster (intente ejecutar ?geom\_raster) para usarlo con los datos de anio, mes y mediana y comenta que observas.
* Agrupe la tabla temperatura usando la columna anio para obtener el promedio de la columna mediana.
* Con lo anterior vuelva a obtener un grafico del promedio obtenido anteriormente por año.
* Si tuviese que escoger alguna de las dos alternativas de gráficos de líneas para presentarla a alguién que nuevo en el tema, ¿Cual escogería y por que?

## Ejercicio 3

Trabajaremos con los siguientes datos utilizados anteriormente:

miembros <- read\_csv("https://goo.gl/Tigfaj")  
glimpse(miembros)

## Observations: 83  
## Variables: 3  
## $ funcion <chr> "analista", "analista", "analista", "analista", "...  
## $ nombre <chr> "Alberto Recarte Garcia Andrade", "Alejandro Couc...  
## $ organizacion <chr> "banco", "consultora", "banco", "banco", "banco",...

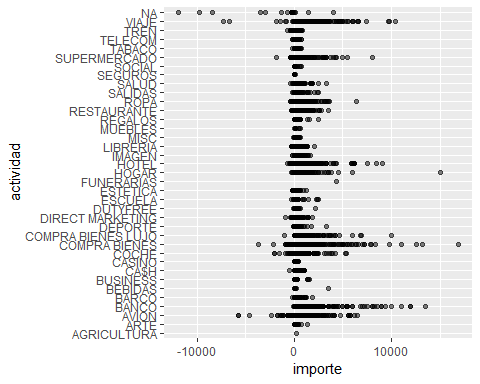
movimientos <- read\_csv("https://goo.gl/fT3jwT")  
glimpse(movimientos)

## Observations: 77,202  
## Variables: 8  
## $ nombre <chr> "Alberto Recarte Garcia Andrade", "Alberto ...  
## $ fecha <chr> "2003-01-03 23:00:00 UTC", "2003-01-03 23:0...  
## $ hora <int> 12, 12, 19, 15, 16, 15, 10, 12, 15, 15, 15,...  
## $ minuto <int> 30, 32, 7, 31, 5, 27, 20, 58, 25, 28, 28, 2...  
## $ importe <dbl> 38.70, 14.60, 95.62, 49.13, 13.94, 80.00, 5...  
## $ comercio <chr> "RCG OFICINA", "MANZANIL AREA", "REST REAL ...  
## $ actividad\_completa <chr> "CONFECCION TEXTIL EN GENERAL", "HOTELES,MO...  
## $ actividad <chr> "ROPA", "HOTEL", "RESTAURANTE", "COCHE", "C...

Con la función ymd\_hms del paquete lubridate puede *parsear* (convertir) la columna fecha (que es texto!) a fecha-hora. Luego con la función as.Date puedes tranformar el campo de fecha-hora a fecha.

* Con lo anterior agrupe por dias y obtenga el número de operaciones, como también el importe total por día de la tabla movimientos.
* Grafique, de forma separada estas dos variables (por fecha). ¿Qué tipo de geom\_ puede ser opción?
* Obtenga la suma de los importes por actividad de los miembros
* Con la tabla anterior grafíquelos usando geom\_col.
* Obtenga la suma de los importes por actividad y hora, luego utilice geom\_tile para mostrar la información obtenida y comente.
* ¿Que es lo que muestra el siguiente gráfico, y que podría hacer con esta información en el caso que usted sea?

library(tidyverse)  
  
ggplot(movimientos) +  
 geom\_point(aes(importe, actividad), alpha = 0.5)



* Usando la imagen anterior identifique usted 3 casos (puntos/movimientos) que le llamen la atentción y descríbalos: Describiendo la hora, quien realizó la operación, el monto, la actividad, etc, y comente si el dato parece digno de investigar o no.