

Análisis de resultados con *R*

PISA, PIAAC, TIMSS y PIRLS

Arturo Caballero Altamirano
aacaballero@outlook.es
caballero000@gmail.com

2 de abril de 2018

Notas elaboradas para el espacio de aprendizaje *Sistematización y aplicación de datos en psicometría y evaluación*.

EL objetivo es brindar herramientas para analizar resultados de evaluaciones internacionales.

El taller se centra en los resultados de PISA aunque las librerías permiten que se pueda trabajar con datos de otras evaluaciones. Las fuentes principales de estas notas son Intsvy [1, Caro, D.] y el reporte técnico de PISA 2015 [6, OECD].

Contenido

Introducción	iii
Requisitos	iii
Comandos básicos en <i>R</i>	iii
Datos	viii
1 Datos generales	1
1.1 Medias por país	1
1.2 Niveles de desempeño	4
2 Relación con otros datos	6
2.1 Producto Interno Bruto per cápita	6
3 Índices incluidos en los datos	10
3.1 Estatus socioeconómico	10
4 Ítems de opción múltiple	12
4.1 Financiamiento	12
5 Presentación de resultados	15
Bibliografía	16

Introducción

Requisitos

Para el taller se necesita tener instalado *R* y *RStudio*, para ello se puede consultar la página de *R-project* (<https://www.r-project.org/>) y de *RStudio* (<https://www.rstudio.com/>). Si es la primera vez que se realizará se recomienda seguir algún tutorial de youtube como [2, Latincoder], [3, Tellez] o [4, Ing. Matemáticos].

Para el desarrollo de las actividades también será necesario contar con las librerías: *intsvy* [1, Caro, D.], *ggplot2* [9, Wickham, H], *foreign* [7, R Core Team], *RColorBrewer* [5, Neuwirth, E.]. Estas se pueden instalar con el comando *install.packages(libreria)* o bien siguiendo los pasos descritos a continuación:

1. Abrir *RStudio* y desplegar el menú superior *Herramientas* (Tools).
2. Presionar *Instalar paquetes* (Install packages).
3. Escribir el nombre de la librería.
4. Presionar *Instalar* (Install).
5. Esperar a que se termine la descarga e instalación.
6. Repetir el proceso para la siguiente librería.

Comandos básicos en *R*

A continuación se describen algunos comandos que son de utilidad para los objetivos del taller, no obstante los conocimientos se pueden ampliar tomando algún curso en línea de forma *gratuita* (o de paga) en Coursera, Datacamp o Youtube.

```
#### Operaciones básicas.  
5 + 5 # Suma.  
  
## [1] 10  
  
6 - 1 # Resta.  
  
## [1] 5  
  
2 * 3 # Multiplicación.  
  
## [1] 6  
  
10 / 2 # División.
```

```
## [1] 5

2**3 # Elevar a una potencia.

## [1] 8

#### Limpiar la consola.
cat("\014") #También se puede hacer de manera manual presionando Ctrl + L.

#### Establecer el directorio (carpeta) en la que se trabajará.
#### Si el sistema operativo es Windows, el directorio luce como C://caballero.
dir <- "/home/caballero/Documentos/Datos/PISA15"

setwd(dir) # Establece el directorio de trabajo.
getwd() # Devuelve el directorio en el que se esta trabajando.

#### Crear listas de elementos.
pais <- c('Honduras','Mexico','Ecuador','Guatemala','Paraguay', NA, NA)
escuelas <- c(12,13,14,15,12,13,NA)
evaluados <- c(21000, NA, 21000,23400, 26800, 25000, 600)
fecha <- as.Date(c('2010-11-1','2008-3-25','2008-3-25','2007-3-14','2008-3-25',
                  NA, NA))

#
#### Crear un data frame a partir de listas de elementos.
datos <- data.frame(pais, escuelas, evaluados, fecha)
datos # Muestra los datos.

##      pais escuelas evaluados      fecha
## 1 Honduras      12     21000 2010-11-01
## 2  Mexico      13         NA 2008-03-25
## 3 Ecuador      14     21000 2008-03-25
## 4 Guatemala     15     23400 2007-03-14
## 5 Paraguay      12     26800 2008-03-25
## 6    <NA>      13     25000    <NA>
## 7    <NA>      NA        600    <NA>

#### Explorar el data frame.
names(datos) # Brinda el nombre de las columnas del data frame.

## [1] "pais"      "escuelas"  "evaluados" "fecha"

summary(datos) # Describe de manera general el data frame.

##      pais      escuelas      evaluados      fecha
## Ecuador :1  Min. :12.00  Min. : 600  Min. :2007-03-14
## Guatemala:1  1st Qu.:12.25  1st Qu.:21000 1st Qu.:2008-03-25
## Honduras :1  Median :13.00  Median :22200 Median :2008-03-25
## Mexico :1  Mean :13.17  Mean :19633  Mean :2008-07-17
## Paraguay :1  3rd Qu.:13.75  3rd Qu.:24600 3rd Qu.:2008-03-25
## NA's :2  Max. :15.00  Max. :26800  Max. :2010-11-01
##      NA's :1  NA's :1  NA's :2
```

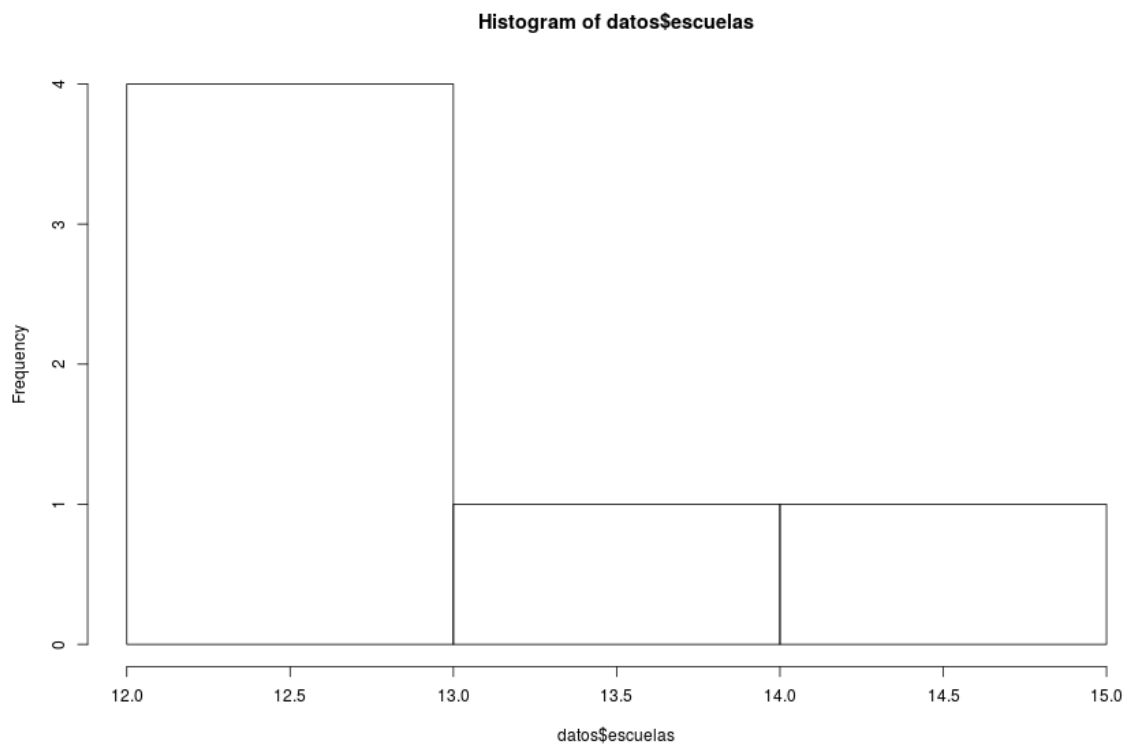
```
class(datos$fecha) # Tipo de dato

## [1] "Date"

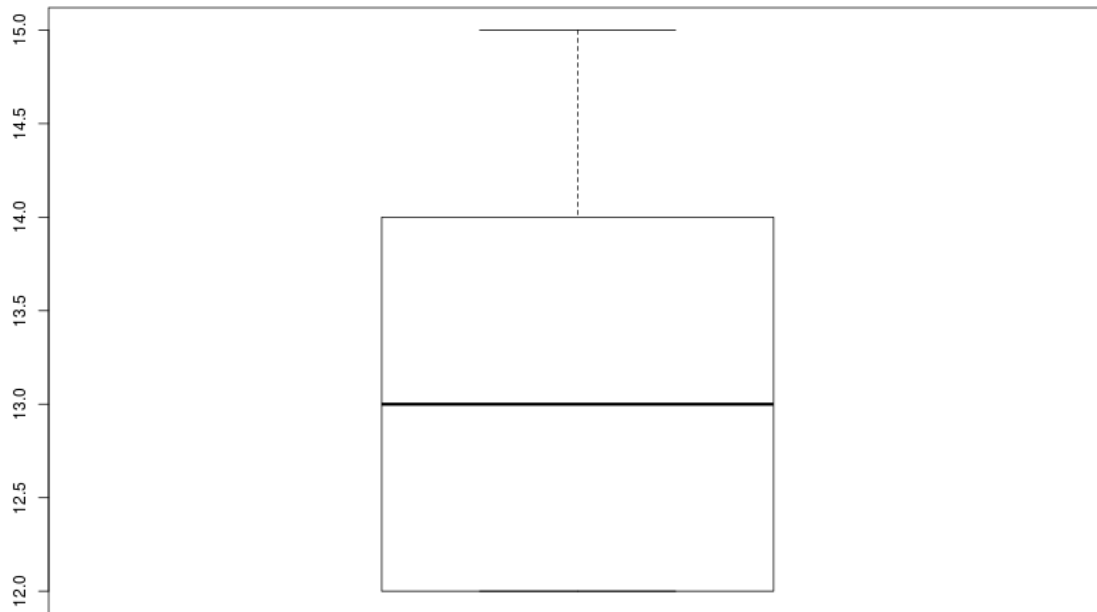
# Media de la columna escuelas
mean(datos$escuelas, na.rm = TRUE)

## [1] 13.16667

# Mediana - median, Desviación - sd, varianza - var.
hist(datos$escuelas) # histograma
```



```
boxplot(datos$escuelas)
```



En *R*, al igual que en otros paquetes estadísticos, se pueden ejecutar procedimientos sin tomar en cuenta las hipótesis que los acompañan por lo que deberá tenerse cuidado al ejecutar los comandos.

```
# Regresion lineal lm(y~x): y = A + Bx
x <- lm(datos$evaluados ~ datos$escuelas)
x

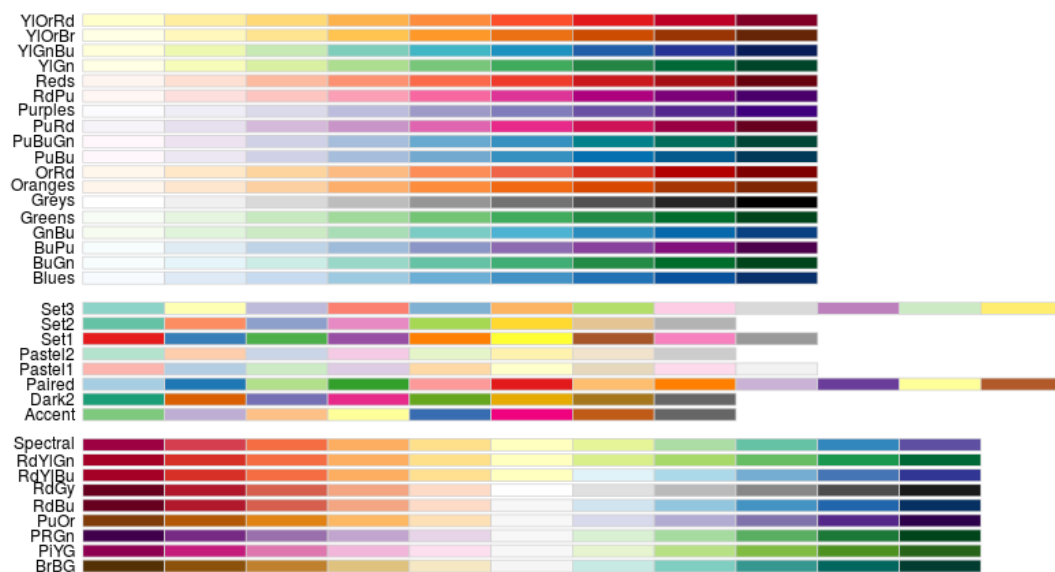
##
## Call:
## lm(formula = datos$evaluados ~ datos$escuelas)
##
## Coefficients:
##      (Intercept)  datos$escuelas
##          30117.6          -505.9

#### Elimina el data frame "datos".
rm(datos)
```

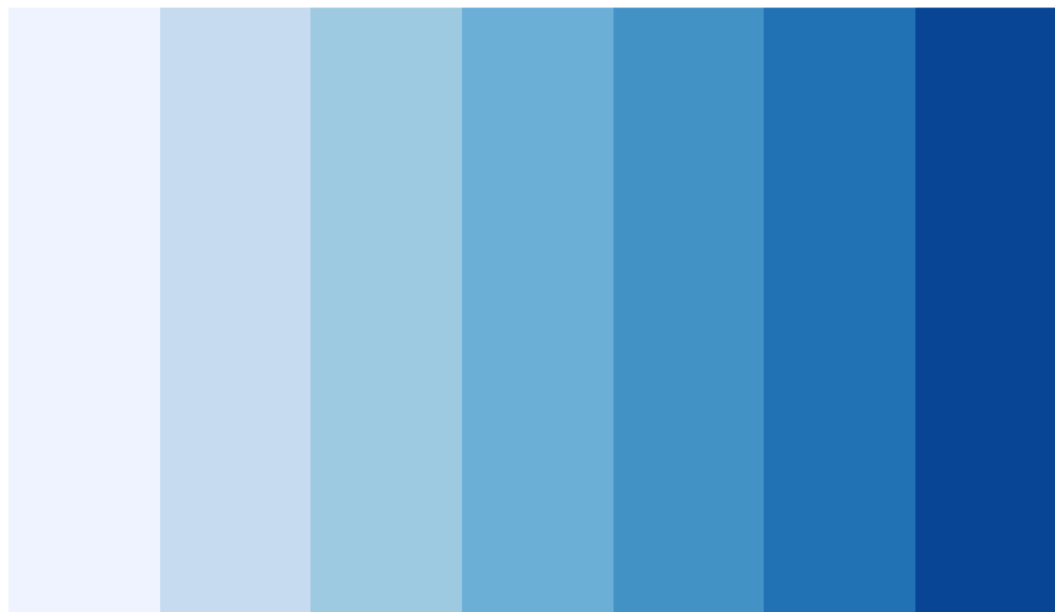
Ejercicio 0.0.1. ¿Con cuántos datos se puede obtener una regresión lineal? ¿Cómo se interpreta?

```
#### Llamar a una librería, ejemplo la librería ggplot2.
library(ggplot2)
```

```
#### Escalas de colores predefinidas.
library("RColorBrewer") # Activa la librería.
display.brewer.all() # Despliega todas las escalas posibles.
```



```
# Muestra una escala en específico con n=7 tonos.
display.brewer.pal(n = 7, name = 'Blues')
```

Blues (sequential)

Ejercicio 0.0.2. Obtener una escala de colores tipo semáforo (rojo, amarillo y verde).

Datos

Los tablas de datos por aplicación pueden ser descargadas de los siguientes enlaces:

Evaluaciones internacionales.

PISA 2015: Programme for International Student Assessment.

<http://www.oecd.org/pisa/data/2015database/>

PISA-D: Programme for International Student Assessment.

<http://www.oecd.org/pisa/pisa-for-development/>

PIAAC: Programme for the International Assessment of Adult Competencies.

<http://www.oecd.org/skills/piaac/publicdataandanalysis/>

TIMSS y PIRLS: Trends in International Mathematics and Science Study (TIMSS) y Progress in International Reading Literacy Study (PIRLS).

<https://timssandpirls.bc.edu/>

Indicadores internacionales y datos para contextualizar los resultados.

Banco Mundial: <https://datos.bancomundial.org/>

Programa de las Naciones Unidas para el Desarrollo: <http://www.un.org/es/databases/index.html>

UNESCO Institute for Statistics: <http://data.uis.unesco.org/>

Indicadores nacionales: Se pueden descargar en los Institutos de estadística, Banco Central, Ministerios o Secretarías.

Una vez descargados los datos de las evaluaciones internacionales, en formato *SPSS*, se pueden cargar a *R* de la siguiente forma:

```
##### Ejemplo con los casos de PISA 2015.
##### El formato .sav es porque los archivos están en SPSS.
library(intsvy) # se carga la librería.

dir <- "/home/caballero/Documentos/Datos/PISA2015"
setwd(dir) # Se establece el directorio.

#### Se cargan los datos.
pisa15 <- pisa.select.merge(folder = dir,
                           student.file="CY6_MS_CMB_STU_QQQ.sav",
                           school.file="CY6_MS_CMB_SCH_QQQ.sav",
                           student= c("ESCS", "ST004D01T", "TEACHSUP",
                                       "JOYSCIE"),
                           school = c("EDUSHORT", "SCHLTYPE"))
```

En caso de que no se quiera ejecutar el merge cada vez que se usan los datos también pueden guardarse en formato *.Rda*.

Ejercicio 0.0.3. Investigar la sentencia necesaria para guardar el data frame en formato *.Rda*.

Ejercicio 0.0.4. Descargar de la página del Banco Mundial los datos del PIB per cápita.

Capítulo 1

Datos generales

1.1 Medias por país

Con el comando *pisa2015.mean.pv* se obtienen las medias de los datos usando los valores plausibles (PV). Si se quiere ahondar en el significado y uso de los PV se puede leer el artículo *What are plausible values and why are they useful?* [8, Von Davier]. A grandes razgos los valores plausibles son valores imputados que pueden ser usados para estimar el desempeño de los estudiantes, comúnmente se generan entre 5 y 10 por evaluado para tratar de mitigar el error de imputación y no son ni suma de aciertos ni una transformación lineal de ellos.

```
library(intsvy) # Se carga la libreria.
medias_scie <- pisa2015.mean.pv(pvlabel = "SCIE",
                               by = "CNT",
                               data = pisa15)

#
medias_math <- pisa2015.mean.pv(pvlabel = "MATH",
                                by = "CNT",
                                data = pisa15)

#
medias_read <- pisa2015.mean.pv(pvlabel = "READ",
                                 by = "CNT",
                                 data = pisa15)
```

```
#### Se muestran solo las primeras filas de la tabla obtenida para Ciencias.
head(medias_scie)

##   CNT  Freq  Mean s.e.    SD  s.e
## 1 ALB  5215 427.22 3.28  78.48 1.45
## 2 ARE 14167 436.73 2.42  99.14 1.06
## 3 AUS 14530 509.99 1.54 102.30 0.92
## 4 AUT  7007 495.04 2.44  97.34 1.31
## 5 BEL  9651 502.00 2.29 100.19 1.24
## 6 BGR  5928 445.77 4.35 101.52 2.10
```

En la tabla que resulta del comando *pisa2015.mean.pv* se pueden ver las columnas: *CNT*, código del país; *Freq*, frecuencia; *Mean*, media; *s.e.*, error estándar de la media; *SD*, desviación estándar y; *s.e.*, error estándar de la desviación.

Muestra información descriptiva de los datos.

summary(medias_scie)

```
##          CNT          Freq          Mean          s.e.
## ALB      : 1   Min.      : 1398   Min.      :331.6   Min.      :1.000
## ARE      : 1   1st Qu.: 5324   1st Qu.:424.6   1st Qu.:2.080
## AUS      : 1   Median   : 5882   Median   :480.6   Median   :2.420
## AUT      : 1   Mean     : 7114   Mean     :467.6   Mean     :2.616
## BEL      : 1   3rd Qu.: 7007   3rd Qu.:505.5   3rd Qu.:2.900
## BGR      : 1   Max.     :32330   Max.     :555.6   Max.     :6.610
## (Other):67
##          SD          s.e
## Min.      : 64.85   Min.      :0.730
## 1st Qu.: 84.38   1st Qu.:1.110
## Median : 90.92   Median   :1.350
## Mean     : 90.17   Mean     :1.423
## 3rd Qu.: 98.74   3rd Qu.:1.580
## Max.     :117.64   Max.     :2.720
##
```

Se muestran solo las primeras filas de la tabla obtenida para Matemáticas.

head(medias_math)

```
##   CNT Freq  Mean s.e.   SD s.e
## 1 ALB 5215 413.16 3.45 86.23 1.62
## 2 ARE 14167 427.48 2.41 96.53 1.29
## 3 AUS 14530 493.90 1.61 93.06 1.23
## 4 AUT 7007 496.74 2.86 95.14 1.83
## 5 BEL 9651 506.98 2.35 97.36 1.47
## 6 BGR 5928 441.19 3.95 97.18 2.37
```

summary(medias_math)

```
##          CNT          Freq          Mean          s.e.
## ALB      : 1   Min.      : 1398   Min.      :327.7   Min.      :1.110
## ARE      : 1   1st Qu.: 5324   1st Qu.:417.2   1st Qu.:2.170
## AUS      : 1   Median   : 5882   Median   :478.6   Median   :2.530
## AUT      : 1   Mean     : 7114   Mean     :461.8   Mean     :2.734
## BEL      : 1   3rd Qu.: 7007   3rd Qu.:501.7   3rd Qu.:3.030
## BGR      : 1   Max.     :32330   Max.     :564.2   Max.     :6.910
## (Other):67
##          SD          s.e
## Min.      : 68.44   Min.      :0.830
## 1st Qu.: 83.73   1st Qu.:1.310
## Median : 88.29   Median   :1.530
## Mean     : 88.46   Mean     :1.636
## 3rd Qu.: 93.89   3rd Qu.:1.830
## Max.     :110.28   Max.     :3.520
##
```

Se muestran solo las primeras filas de la tabla obtenida para Lectura.

```
head(medias_read)
```

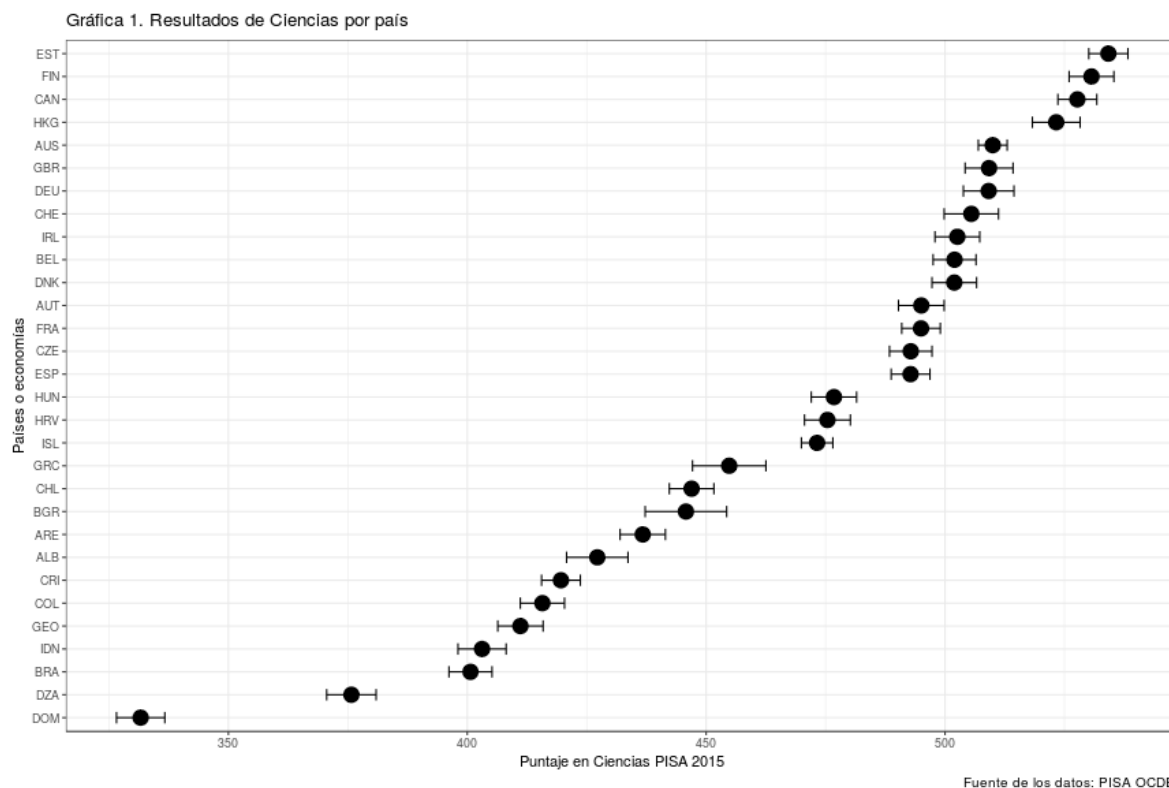
```
##   CNT  Freq  Mean s.e.    SD  s.e
## 1 ALB  5215 405.26 4.13  96.63 1.85
## 2 ARE 14167 433.54 2.87 105.73 1.39
## 3 AUS 14530 502.90 1.69 102.69 1.12
## 4 AUT  7007 484.87 2.84 101.14 1.54
## 5 BEL  9651 498.52 2.42 100.22 1.52
## 6 BGR  5928 431.72 5.00 114.60 2.60
```

```
summary(medias_read)
```

```
##           CNT           Freq           Mean           s.e.
## ALB      : 1   Min.      : 1398   Min.      :346.6   Min.      :1.020
## ARE      : 1   1st Qu.: 5324   1st Qu.:426.9   1st Qu.:2.410
## AUS      : 1   Median : 5882   Median :484.8   Median :2.690
## AUT      : 1   Mean    : 7114   Mean    :463.0   Mean    :2.918
## BEL      : 1   3rd Qu.: 7007   3rd Qu.:499.9   3rd Qu.:3.080
## BGR      : 1   Max.    :32330   Max.    :535.1   Max.    :7.190
## (Other):67
##           SD           s.e
## Min.      : 72.60   Min.      :0.860
## 1st Qu.: 88.14   1st Qu.:1.390
## Median : 95.06   Median :1.580
## Mean      : 94.80   Mean      :1.658
## 3rd Qu.:100.22   3rd Qu.:1.780
## Max.      :120.55   Max.      :3.380
##
```

Gráfica de medias para Ciencias.

```
q <- plot(medias_scie[1:30,], sort = TRUE) # Se grafican los primeros 30 datos.
q + labs(x = "Países o economías", y = "Puntaje en Ciencias PISA 2015",
         title = "Gráfica 1. Resultados de Ciencias por país",
         caption = "Fuente de los datos: PISA OCDE.")
```



Ejercicio 1.1.1. Replicar las gráficas para Matemáticas y Lectura.

Ejercicio 1.1.2. Tomando como ejemplo la gráfica de los resultados de Ciencias, replicar el ejercicio usando únicamente los países latinoamericanos.

1.2 Niveles de desempeño

Los puntos de corte para los niveles de desempeño de las evaluaciones pueden ser consultados en sus páginas respectivas. En el caso de PISA 2015 consultar el informe técnico [6, OECD].

```
#### Tabla con los niveles de desempeño por país.
df <- pisa2015.ben.pv(pvlabel="SCIE",
  cutoff = c(260.54, 334.94, 409.54, 558.73, 633.33, 707.93),
  by="CNT",
  data=pisa15)

names(df) # Nombre de las columnas

## [1] "CNT" "Benchmarks" "Percentage" "Std. err."

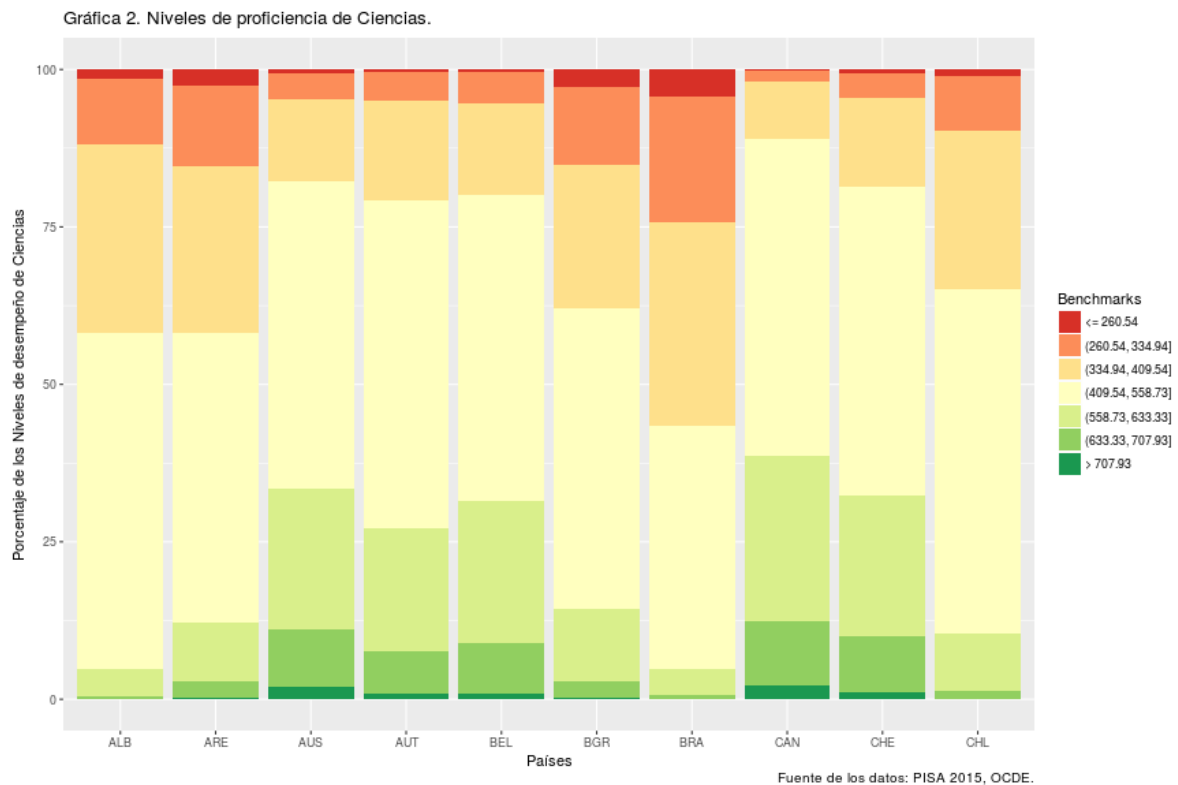
df[1:7,] # Primeros 7 datos de la tabla df.

## CNT Benchmarks Percentage Std. err.
## 1 ALB <= 260.54 1.55 0.31
## 2 ALB (260.54, 334.94] 10.34 0.76
## 3 ALB (334.94, 409.54] 29.88 1.22
## 4 ALB (409.54, 558.73] 53.37 1.39
```

```
## 5 ALB (558.73, 633.33]      4.44      0.56
## 6 ALB (633.33, 707.93]      0.42      0.14
## 7 ALB      > 707.93         0.02      0.04
```

Ejercicio 1.2.1. Describir con detalle que significan las columnas de la tabla *df*. ¿Por qué la tabla se presenta de esta forma?

```
#### Gráfica con los niveles de desempeño.
ggplot(data=df[1:70,],
       aes(x=CNT, y=Percentage, fill=Benchmarks)) +
  geom_bar(stat="identity") +
  scale_fill_manual(values=brewer.pal(n = 7, name = "RdYlGn")) +
  labs(x = "Países",
       y = "Porcentaje de los Niveles de desempeño de Ciencias",
       title = "Gráfica 2. Niveles de proficiencia de Ciencias.",
       caption = "Fuente de los datos: PISA 2015, OCDE.")
```



Ejercicio 1.2.2. ¿Cuántos países se muestran con los primeros 70 datos de la tabla de niveles de desempeño? ¿Cuántos datos tiene la tabla?

Ejercicio 1.2.3. Modificar la sintaxis para que aparezcan los porcentajes dentro de cada columna.

Ejercicio 1.2.4. Replicar la gráfica para Matemáticas y Lectura.

Capítulo 2

Relación con otros datos

2.1 Producto Interno Bruto per cápita

Los datos de las evaluaciones internacionales son una fuente muy importante de información que en conjunto con otras mediciones o indicadores pueden ayudar a entender la situación de la educación en un país, región o grupo de naciones que comparten algún tipo de características. Por ejemplo se pueden comparar con indicadores macro como el Producto Interno Bruto (PIB), el Índice de Desarrollo Humano, el porcentaje del PIB destinado a la educación o el Índice de GINI.

```
#### Importar a R los datos que se encuentran en un archivo .csv.
#### La fuente de los datos es el Banco Mundial.
#### Los datos son del PIB per cápita a precios actuales, fueron descargados el
#### 31 de Marzo de 2018.
pibpc <- read.csv(
  file="/home/caballero/Documentos/Datos/indicadores/pib_percapita_20180331.csv",
  header=TRUE,
  sep=";")
#
names(pibpc) # Muestra los nombres de las columnas que tiene la tabla de datos.

## [1] "Country.Name" "Country.Code" "Indicator.Name" "Indicator.Code"
## [5] "X2000" "X2001" "X2002" "X2003"
## [9] "X2004" "X2005" "X2006" "X2007"
## [13] "X2008" "X2009" "X2010" "X2011"
## [17] "X2012" "X2013" "X2014" "X2015"
## [21] "X2016"

names(pibpc)[1] = "Pais" # Cambia el nombre de la primera columna a Pais.

#### Realiza un subconjunto de los datos con las columnas 1,2 y 20.
pibpc <- pibpc[,c(1,2,3,20)]

head(pibpc) # Muestra los primeros datos de la tabla.

## Pais Country.Code Indicator.Name
## 1 Aruba ABW PIB per cápita (US$ a precios actuales)
## 2 Afganistán AFG PIB per cápita (US$ a precios actuales)
```



```
## 3      Angola      AGO PIB per cápita (US$ a precios actuales)
## 4      Albania     ALB PIB per cápita (US$ a precios actuales)
## 5      Andorra     AND PIB per cápita (US$ a precios actuales)
## 6 El mundo árabe  ARB PIB per cápita (US$ a precios actuales)
##      X2015
## 1      NA
## 2      569.5779
## 3      3695.7937
## 4      3934.8954
## 5      36038.2676
## 6      6435.5255

summary(pibpc) # Estadísticas básicas de los datos.

##                               Pais      Country.Code
## Afganistán                   : 1  ABW      : 1
## África al sur del Sahara      : 1  AFG      : 1
## África al sur del Sahara (BIRF y la AIF) : 1  AGO      : 1
## África al sur del Sahara (excluido altos ingresos): 1  ALB      : 1
## Albania                       : 1  AND      : 1
## Alemania                     : 1  ARB      : 1
## (Other)                       :258 (Other):258
##                               Indicator.Name      X2015
## PIB per cápita (US$ a precios actuales):264  Min.   : 300.7
##                                              1st Qu.: 1907.1
##                                              Median : 5882.0
##                                              Mean   : 13969.4
##                                              3rd Qu.: 16130.0
##                                              Max.   :168146.0
##                                              NA's   :20
```

Para representar los datos del PIB per cápita en conjunto con los resultados de PISA se deben unir las dos tablas, para ello se puede usar el comando *merge*.

```
names(medias_scie)

## [1] "CNT" "Freq" "Mean" "s.e." "SD" "s.e"

names(pibpc)

## [1] "Pais" "Country.Code" "Indicator.Name" "X2015"

pisa_scie <- merge(medias_scie,
                  pibpc,
                  by.x = "CNT", # Columna en la tabla medias_scie (medias_scie).
                  by.y = "Country.Code") # Columna en la tabla del PIB (pibpc).

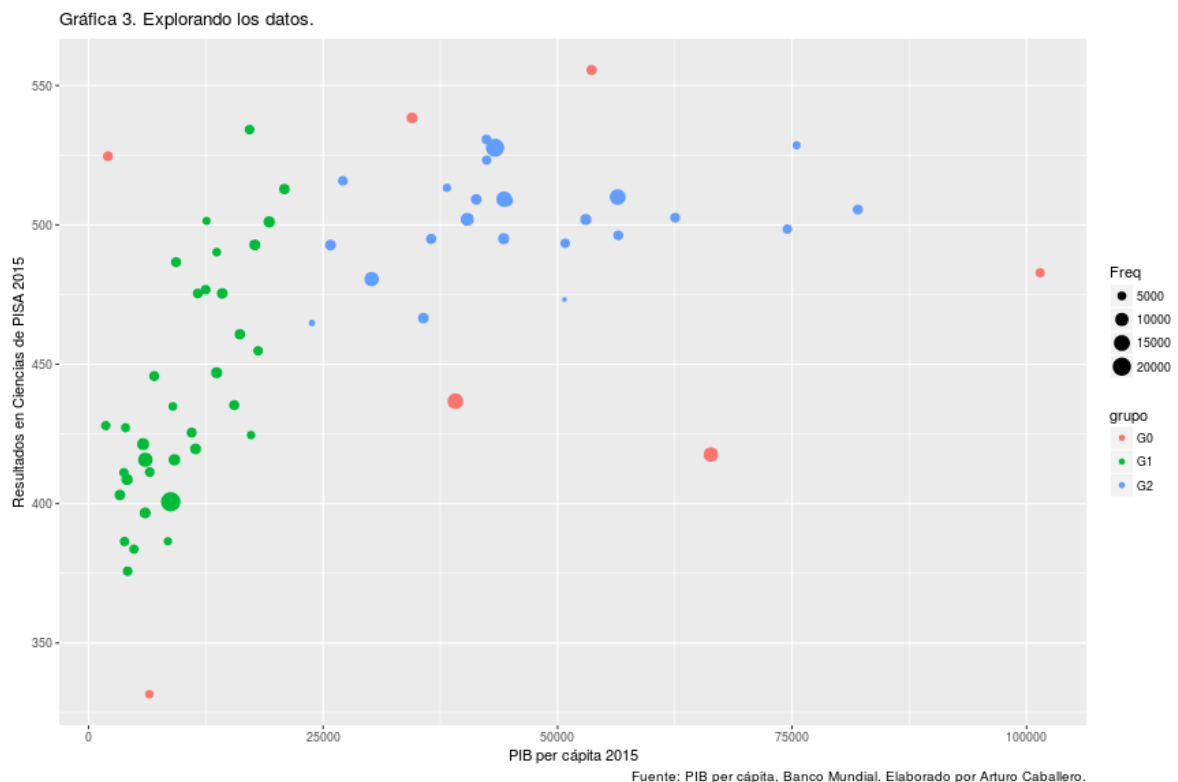
names(pisa_scie)

## [1] "CNT" "Freq" "Mean" "s.e."
## [5] "SD" "s.e" "Pais" "Indicator.Name"
## [9] "X2015"
```

Ejercicio 2.1.1. ¿Qué significa el PIB per cápita? ¿Por qué sería interesante cruzarlo con los datos de PISA?

```
#### Con esta sintaxis solo se prepara la tabla para mostrar los resultados
#### por grupos.
pisa_scie$grupo <- "G2"
pisa_scie$grupo[pisa_scie$X2015 < 23000] <- "G1"
pisa_scie$grupo[pisa_scie$CNT %in% c("VNM", "ARE", "DOM", "QAT", "JPN", "SGP", "LUX")] <- "G0"

##### Gráfica de los datos.
ggplot(pisa_scie, aes(x=X2015, y=Mean)) +
  geom_jitter(aes(col=grupo, size=Freq)) +
  labs(x = "PIB per cápita 2015",
       y = "Resultados en Ciencias de PISA 2015",
       title = "Gráfica 3. Explorando los datos.",
       caption = "Fuente: PIB per cápita, Banco Mundial. Elaborado por Arturo Caballero.")
```



```
#### Se obtiene el subconjunto del grupo G1.
pisa_scie1 <- pisa_scie[pisa_scie$grupo == "G1",]
```

```
##### Gráfica del grupo G1.
theme_set(theme_bw()) # Se establece el tema que se ocupará.

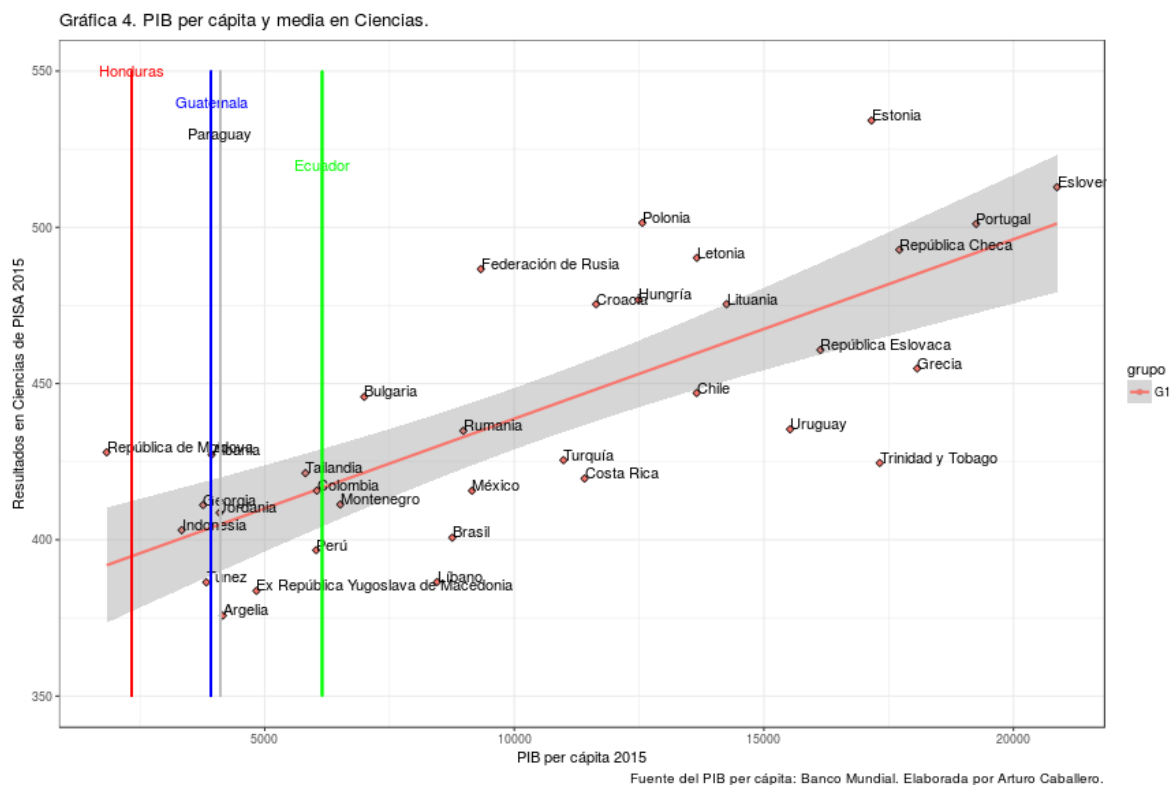
q <- ggplot(pisa_scie1, aes(x=X2015, y=Mean)) + # Gráfica de dispersión.
  geom_jitter(aes(col=grupo)) +
```

```

geom_smooth(aes(col=grupo),method="lm",se=TRUE) +
geom_point(shape=5) +
geom_text(aes(label=Pais),hjust=0, vjust=0) +
geom_segment(aes(x = 2326.156,y = 550, xend = 2326.156, yend = 350),
  colour = "red",show.legend = FALSE) +
  annotate("text",x = 2326.156,y = 550,label = "Honduras", color = "red") +
geom_segment(aes(x = 3923.573,y = 550, xend = 3923.573, yend = 350),
  colour = "blue",show.legend = FALSE) +
  annotate("text",x = 3923.573,y = 540,label = "Guatemala",color = "blue") +
geom_segment(aes(x = 4109.368,y = 550, xend = 4109.368, yend = 350),
  colour = "grey",show.legend = FALSE) +
  annotate("text",x = 4109.368,y = 530,label = "Paraguay",color = "black") +
geom_segment(aes(x = 6150.156,y = 550, xend = 6150.156, yend = 350),
  colour = "green",show.legend = FALSE) +
  annotate("text",x = 6150.156,y = 520,label = "Ecuador",color = "green")

q + labs(x = "PIB per cápita 2015", y = "Resultados en Ciencias de PISA 2015",
  title = "Gráfica 4. PIB per cápita y media en Ciencias.",
  caption = "Fuente del PIB per cápita: Banco Mundial. Elaborada por Arturo Caballero.")

```



Ejercicio 2.1.2. ¿Es necesario que aparezca la zona sombreada? ¿Qué nos dice? ¿Qué métodos de agrupación conoces?

Ejercicio 2.1.3. Replicar la gráfica del PIB per cápita y Ciencias usando, en lugar del PIB per cápita, el Índice de Desarrollo Humano.

Ejercicio 2.1.4. Replicar la gráfica del PIB per cápita y Ciencias usando, en lugar del PIB per cápita, un indicador que se considere relevante para el país.

Capítulo 3

Índices incluidos en los datos

3.1 Estatus socioeconómico

La mayoría de las evaluaciones internacionales incluyen en sus tablas de datos índices como el del Estatus Socioeconómico, entre otros. Una forma de trabajar con ellos es encontrar la media por país y usarla para contextualizar los resultados.

```
#### Estatus socioeconómico medio por país.
media_escs <- pisa2015.mean(variable = "ESCS", by = "CNT" , data = pisa15)

#### **** Cuando se trabajen con los datos aparecerán advertencias ****
#               porque el renglón de ALB tiene datos vacios (NaN).

#### Agregar las medias de PISA de Ciencias.
media_escs <- merge(medias_scie,
                    media_escs,
                    by.x = "CNT", # columna de la tabla medias_scie
                    by.y = "CNT") # columna de la tabla media_escs
#
head(media_escs)

##   CNT Freq.x Mean.x s.e..x   SD.x s.e.x Freq.y Mean.y s.e..y SD.y s.e.y
## 1 ALB   5215 427.22   3.28  78.48  1.45     0   NaN   NaN 0.00  0.00
## 2 ARE  14167 436.73   2.42  99.14  1.06 13869   0.50   0.01 0.73  0.01
## 3 AUS  14530 509.99   1.54 102.30  0.92 13989   0.27   0.01 0.78  0.01
## 4 AUT   7007 495.04   2.44  97.34  1.31  6939   0.09   0.02 0.85  0.01
## 5 BEL   9651 502.00   2.29 100.19  1.24  9452   0.16   0.02 0.90  0.01
## 6 BGR   5928 445.77   4.35 101.52  2.10  5793  -0.08   0.03 0.98  0.02

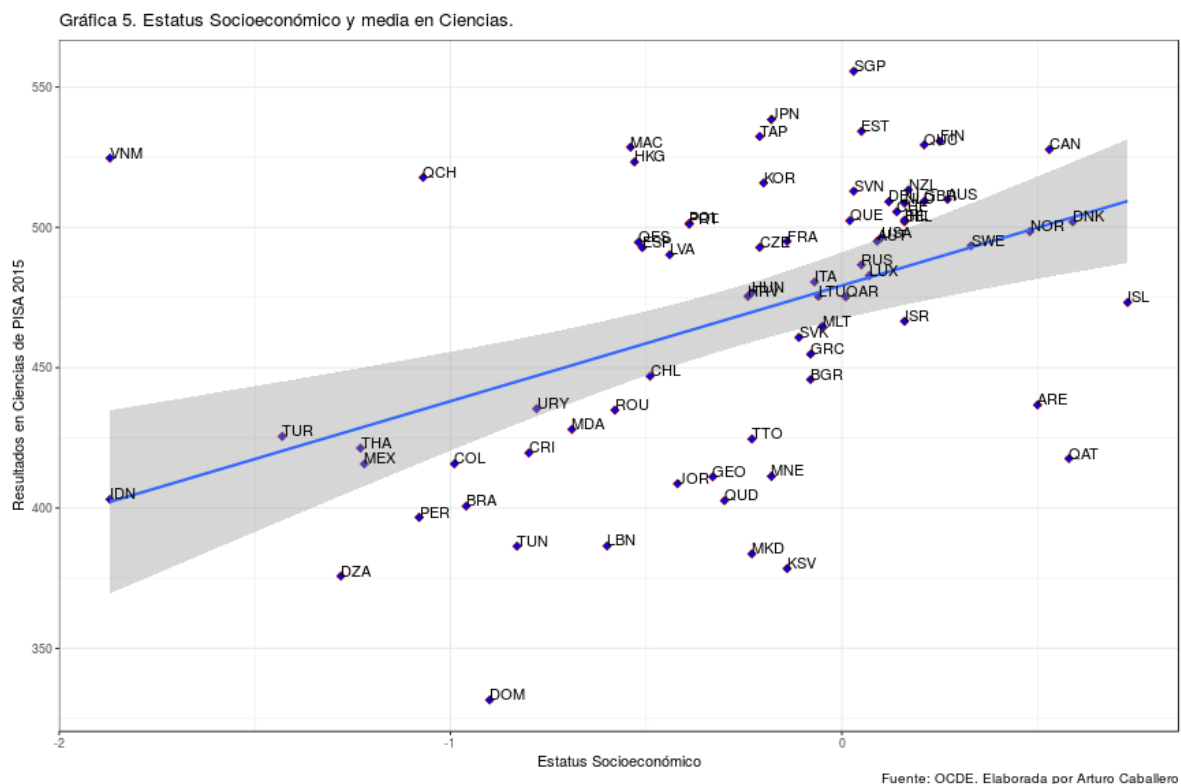
names(media_escs)[3] = "Ciencias" # Editar el nombre de los campos.
#
names(media_escs)[8] = "ESCS" # Editar el nombre de los campos.
#
names(media_escs)

## [1] "CNT"      "Freq.x"   "Ciencias" "s.e..x"   "SD.x"     "s.e.x"
## [7] "Freq.y"   "ESCS"     "s.e..y"   "SD.y"     "s.e.y"
```

Ejercicio 3.1.1. ¿Por qué después de ejecutar el *merge* hay nombres de las columnas que parecen repetirse? ¿Por qué el nombre de las columnas termina en *.x* o *.y*?

```
#### Gráfica de la media del Estatus socioeconómico en
#### conjunto con los resultados de PISA.
q <- ggplot(media_escs, aes(x=ESCS, y=Ciencias)) +
  geom_point(shape=23,
             fill="blue",
             color="darkred",
             size=2) +
  geom_smooth(method=lm, se=TRUE) + # Regresión lineal
  geom_text(aes(label=CNT), hjust=0, vjust=0)
#
q + labs(x = "Estatus Socioeconómico",
        y = "Resultados en Ciencias de PISA 2015",
        title = "Gráfica 5. Estatus Socioeconómico y media en Ciencias.",
        caption = "Fuente: OCDE. Elaborada por Arturo Caballero")

## Warning: Removed 1 rows containing non-finite values (stat_smooth).
## Warning: Removed 1 rows containing missing values (geom_point).
## Warning: Removed 1 rows containing missing values (geom_text).
```



Ejercicio 3.1.2. ¿Qué significan las advertencias (*Warnings*) que salen después de realizar la gráfica?

Ejercicio 3.1.3. Buscar en el *Codebook de PISA 2015* que significa el índice *JOYSCIE* y posteriormente replicar la gráfica usando las medias de Ciencias y el índice *JOYSCIE*.

Capítulo 4

Ítems de opción múltiple

Las evaluaciones regularmente incluyen instrumentos que exploran el contexto social, económico, laboral, cultural o familiar del evaluado a través de ítems de opción múltiple, por lo que es importante poder realizar análisis con este tipo de variables.

4.1 Financiamiento

La variable financiamiento, SCHLTYPE, tiene tres posibles opciones de respuesta y un código para valores perdidos:

- 1 Privado (Private independent).
- 2 Mixto (Private Government - dependent).
- 3 Público (Public).
- 9 NA.

```
# Medias de Ciencias por país y financiamiento.
sost <- pisa2015.mean.pv(pvlabel = "SCIE", by=c("CNT", "SCHLTYPE"), data = pisa15)
names(sost)

## [1] "CNT"      "SCHLTYPE" "Freq"      "Mean"      "s.e."      "SD"
## [7] "s.e"

head(sost)

##   CNT SCHLTYPE Freq   Mean  s.e.   SD  s.e
## 1 ALB        1  422 479.97 11.93 74.98 4.57
## 2 ALB        3 4708 421.11  2.72 76.86 1.35
## 3 ALB       <NA>   85 439.43 15.17 66.38 4.90
## 4 ARE        1 5706 473.94  3.60 99.55 1.63
## 5 ARE        2  149 362.99 11.78 65.42 5.33
## 6 ARE        3 5158 394.41  2.83 82.41 1.41
```

Ejercicio 4.1.1. Obtener las medias de Ciencias por país y por sexo.

Ejercicio 4.1.2. ¿Es posible hacer regresiones con cualquier tipo de variable (nominal, ordinal, categórica)?

```
#### Regresión con tipo de escuela por país.
reg_sost <- pisa2015.reg.pv(pvlabel="SCIE",
                           x="SCHLTYPE" ,
                           by = "CNT",
                           data=pisa15)

#### Seleccionar un país en específico.
reg_sost$ALB$reg # resultados de Albania (ALB).

##              Estimate Std. Error   t value
## (Intercept) 509.39604081 18.46220259 27.591293
## SCHLTYPE    -29.42796430  6.25854890 -4.702043
## R-squared    0.05025615  0.02503946  2.007078

reg_sost$CHL$reg # resultados de Chile (CHL).

##              Estimate Std. Error   t value
## (Intercept) 541.9326190  8.88917227  60.965476
## SCHLTYPE    -41.9213927  3.74996277 -11.179149
## R-squared    0.0978163  0.01633806   5.987019

reg_sost$COL$reg # resultados de Colombia (COL).

##              Estimate Std. Error   t value
## (Intercept) 501.1399902 10.13781294  49.432752
## SCHLTYPE    -32.0502722  3.69336589 -8.677795
## R-squared    0.1003581  0.02240493   4.479285

reg_sost$URY$reg # resultados de Uruguay (URY).

##              Estimate Std. Error   t value
## (Intercept) 550.9430629  7.12684470  77.305327
## SCHLTYPE    -42.9238614  2.64883510 -16.204807
## R-squared    0.1279769  0.01586348   8.067393

#### Agregamos el índice ESCS.
reg_sost_escs <- pisa2015.reg.pv(pvlabel="SCIE",
                                 x=c("SCHLTYPE", "ESCS"),
                                 by = "CNT",
                                 data=pisa15)

reg_sost_escs$DOM$reg # resultados de República Dominicana (DOM).

##              Estimate Std. Error   t value
## (Intercept) 397.8151662 13.41374301  29.657283
## SCHLTYPE    -18.2667784  5.02457648 -3.635486
## ESCS         20.2062561  2.68998494  7.511661
## R-squared    0.1587171  0.02836855   5.594827

reg_sost_escs$CHL$reg # resultados de Chile (CHL).

##              Estimate Std. Error   t value
```

```
## (Intercept) 511.7442380 8.95829711 57.125169
## SCHLTYPE    -22.7030878 3.97973673 -5.704671
## ESCS        26.0515200 1.74202002 14.954776
## R-squared    0.1841447 0.01573215 11.704997

reg_sost_escs$COL$reg  # resultados de Colombia (COL).

##              Estimate Std. Error  t value
## (Intercept) 489.5361880 7.90106277 61.95827
## SCHLTYPE    -20.0610216 3.10212238 -6.46687
## ESCS        19.7856921 1.69744011 11.65619
## R-squared    0.1638286 0.02395088  6.84019

reg_sost_escs$URY$reg  # resultados de Uruguay (URY).

##              Estimate Std. Error  t value
## (Intercept) 523.2255618 5.87117847 89.11764
## SCHLTYPE    -25.7217056 2.48094599 -10.36770
## ESCS        23.5633376 1.54431589 15.25811
## R-squared    0.1965605 0.01547774 12.69956
```

Ejercicio 4.1.3. ¿Cómo se lee cada regresión?

Ejercicio 4.1.4. Sacar la regresión por país con la variable sexo.

Capítulo 5

Presentación de resultados

Proyecto 5.0.1. Seleccionar un público objetivo y preparar una presentación, informe o tríptico con los resultados obtenidos durante el taller.

Al preparar la presentación tomar en cuenta los siguientes puntos:

1. Identificar los principales grupos de interés y hacer productos adecuados para cada uno.
2. Identificar el mensaje o información que se quiere transmitir, para que esto se haga de manera clara a través de todos los productos.
3. Buscar los medios de comunicación adecuados que sirvan para llegar a cada grupo de interés.
4. Revizar que materiales ya existen y usarlos, no es necesario duplicar esfuerzos.
5. En caso de que haya materiales que contextualizar hay que solicitar permiso y pedir consejo sobre los objetivos que se tenía cuando se generaron.
6. Poner al alcance de todos las presentaciones, publicaciones, datos y sintaxis utilizadas.
7. Todos los productos deben elaborarse con transparencia y de preferencia estar acompañados de un informe técnico que explique como se realizaron.

Bibliografía

- [1] Daniel H. Caro and Przemysław Biecek. *Intsvy: An R Package for Analyzing International Large-Scale Assessment Data*, volume 81. 2017.
- [2] Latincoder. RStudio, instalar y configurar tu entorno de desarrollo - Tutorial R. *Archivo de video. Recuperado de <https://www.youtube.com/watch?v=M93KUICRwbk>*, 2014.
- [3] Carlos Tellez Martinez. Instalación de programa R y RStudio. *Archivo de video. Recuperado de <https://www.youtube.com/watch?v=mQ1ySyFVttk>*, 2015.
- [4] Ingenieros Matemáticos. Como instalar R y RStudio - windows 7, 8, 10. *Archivo de video. Recuperado de <https://www.youtube.com/watch?v=1WXgaa2Spp0>*, 2017.
- [5] Erich Neuwirth. *RColorBrewer: ColorBrewer Palettes*, 2014. R package version 1.1-2.
- [6] OECD. *PISA 2015 Technical Report*. 2017.
- [7] R Core Team. *foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ..., 2017*. R package version 0.8-69.
- [8] Matthias Von Davier, Eugenio Gonzalez, and Robert Mislevy. What are plausible values and why are they useful? *IERI monograph series*, 2:9–36, 2009.
- [9] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.