

Taller 4

Nathalie Bareño, Karen Vargas y Sebastián Rincón

16 de mayo

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.0.5      v dplyr  1.0.3
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(readxl)
library(scales)

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##   discard

## The following object is masked from 'package:readr':
##
##   col_factor

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.0.5

library(stargazer)

##
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.0.5
```

```
## Registered S3 method overwritten by 'GGally':  
##   method from  
##   +.gg   ggplot2
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.0.5
```

Puntos del 1 al 4

Carga de bases de datos

```
NorMen <- read_excel("datos/iabbd_8010_v1 (2).xls",  
                     sheet = "NOR-Men")
```

```
## New names:  
## * ' ' -> ...3  
## * ' ' -> ...4  
## * ' ' -> ...5  
## * ' ' -> ...7  
## * ' ' -> ...8  
## * ...
```

```
NorWom <- read_excel("datos/iabbd_8010_v1 (2).xls",  
                     sheet = "NOR-Women")
```

```
## New names:  
## * ' ' -> ...3  
## * ' ' -> ...4  
## * ' ' -> ...5  
## * ' ' -> ...7  
## * ' ' -> ...8  
## * ...
```

```
USMen <- read_excel("datos/iabbd_8010_v1 (2).xls",  
                    sheet = "USA-Men")
```

```
## New names:  
## * ' ' -> ...3  
## * ' ' -> ...4  
## * ' ' -> ...5  
## * ' ' -> ...7  
## * ' ' -> ...8  
## * ...
```

```
USWom <- read_excel("datos/iabbd_8010_v1 (2).xls",
                    sheet = "USA-Women")
```

```
## New names:
## * ' ' -> ...3
## * ' ' -> ...4
## * ' ' -> ...5
## * ' ' -> ...7
## * ' ' -> ...8
## * ...
```

```
HDI <- read_csv("datos/Human Development Index (HDI) (1).csv",
                skip = 5)
```

```
## Warning: Missing column names filled in: 'X4' [4], 'X6' [6], 'X8' [8],
## 'X10' [10], 'X12' [12], 'X14' [14], 'X16' [16], 'X18' [18], 'X20' [20],
## 'X22' [22], 'X24' [24], 'X26' [26], 'X28' [28], 'X30' [30], 'X32' [32],
## 'X34' [34], 'X36' [36], 'X38' [38], 'X40' [40], 'X42' [42], 'X44' [44],
## 'X46' [46], 'X48' [48], 'X50' [50], 'X52' [52], 'X54' [54], 'X56' [56],
## 'X58' [58], 'X60' [60], 'X62' [62]
```

```
##
## -- Column specification -----
## cols(
##   .default = col_character(),
##   'HDI Rank' = col_double(),
##   X4 = col_logical(),
##   X6 = col_logical(),
##   X8 = col_logical(),
##   X10 = col_logical(),
##   X12 = col_logical(),
##   X14 = col_logical(),
##   X16 = col_logical(),
##   X18 = col_logical(),
##   X20 = col_logical(),
##   X22 = col_logical(),
##   X24 = col_logical(),
##   X26 = col_logical(),
##   X28 = col_logical(),
##   X30 = col_logical(),
##   X32 = col_logical(),
##   X34 = col_logical(),
##   X36 = col_logical(),
##   X38 = col_logical(),
##   X40 = col_logical()
##   # ... with 11 more columns
## )
## i Use 'spec()' for the full column specifications.
```

```
## Warning: 1 parsing failure.
## row col expected actual file
## 207 -- 62 columns 1 columns 'datos/Human Development Index (HDI) (1).csv'
```

```
CPI <- read_delim("datos/data_csv.csv", delim = ",")
```

```
##
## -- Column specification -----
## cols(
##   Jurisdiction = col_character(),
##   '1998' = col_character(),
##   '1999' = col_character(),
##   '2000' = col_character(),
##   '2001' = col_character(),
##   '2002' = col_character(),
##   '2003' = col_character(),
##   '2004' = col_character(),
##   '2005' = col_character(),
##   '2006' = col_character(),
##   '2007' = col_character(),
##   '2008' = col_character(),
##   '2009' = col_character(),
##   '2010' = col_character(),
##   '2011' = col_character(),
##   '2012' = col_character(),
##   '2013' = col_character(),
##   '2014' = col_character(),
##   '2015' = col_character()
## )
```

```
Unem <- read_excel("datos/API_SL.UEM.TOTL.ZS_DS2_es_excel_v2_2253098.xls",
  sheet = "Data")
```

```
## New names:
## * ' ' -> ...3
## * ' ' -> ...4
## * ' ' -> ...5
## * ' ' -> ...6
## * ' ' -> ...7
## * ...
```

Modificación de bases de datos

```
# Índice de Percepción de Corrupción
```

```
CPI <- CPI %>% select(c("Jurisdiction", "2010"))
```

```
CPI <- rename(CPI, Country = "Jurisdiction", Corruption = "2010")
```

```
# Índice de Desarrollo Humano
```

```
HDI <- HDI %>% select(c("Country", "2010"))
```

```
HDI <- rename(HDI, Development = "2010")
```

```

# Desempleo

Unem <- Unem %>% select(c("Data Source", "...55"))

Unem <- rename(Unem, Country = "Data Source", Unemployment = "...55")

Unem <- Unem[-c(1,2,3),]

# Noruega Hombres

NorMen <- NorMen %>% select(c("Norway", "...29"))

NorMen <- rename(NorMen, Country = "Norway", HighMenNor = "...29")

NorMen <- NorMen[-c(1,2,3),]

# Noruega Mujeres

NorWom <- NorWom %>% select(c("Norway", "...29"))

NorWom <- rename(NorWom, Country = "Norway", HighWomNor = "...29")

NorWom <- NorWom[-c(1,2,3),]

# Estados Unidos Hombres

USMen <- USMen %>% select(c("United States", "...29"))

USMen <- rename(USMen, Country = "United States", HighMenUS = "...29")

USMen <- USMen[-c(1,2,3),]

# Estados Unidos Mujeres

USWom <- USWom %>% select(c("United States", "...29"))

USWom <- rename(USWom, Country = "United States", HighWomUS = "...29")

USWom <- USWom[-c(1,2,3),]

```

Unión de bases de datos

```

Union1 <- left_join(NorMen, NorWom, by = "Country")

Union2 <- left_join(USWom, USMen, by = "Country")

UnionY <- left_join(Union1, Union2, by = "Country")

Union3 <- left_join(CPI, HDI, by = "Country")

```

```
UnionX <- left_join(Union3, Unem, by = "Country")

Base <- left_join(UnionY, UnionX, by = "Country")
```

Identificación del tipo de datos

```
sapply(Base, class)
```

```
##      Country  HighMenNor  HighWomNor  HighWomUS  HighMenUS  Corruption
## "character" "character" "character" "character" "character" "character"
## Development Unemployment
## "character" "character"
```

Modificación del tipo de datos

```
base <- as.data.frame(sapply(Base[,2:8], as.numeric))
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
Base <- cbind(Base[,1], base)
```

```
sapply(Base, class)
```

```
##      Country  HighMenNor  HighWomNor  HighWomUS  HighMenUS  Corruption
## "character"  "numeric"    "numeric"  "numeric"  "numeric"  "numeric"
## Development Unemployment
##  "numeric"    "numeric"
```

Punto 5

Usando la función `stargazer()` se realizará una tabla de estadísticas descriptivas para las variables que se incluyen en el estudio. Se incluye esta tabla en el RMarkdown con la misma función.

Para simplificar la visualización de los datos, se tendrán en cuenta tanto en cuadros como gráficos la siguiente nomenclatura:

- NOR-Hombres: cantidad de inmigrantes masculinos altamente cualificados que ingresaron a Noruega en el año 2010.
- NOR-Mujeres: cantidad de inmigrantes femeninos altamente cualificados que ingresaron a Noruega en el año 2010.
- US-Mujeres: cantidad de inmigrantes femeninos altamente cualificados que ingresaron a Estados Unidos en el año 2010.

- US-Hombres: cantidad de inmigrantes masculinos altamente cualificados que ingresaron a Estados Unidos en el año 2010.

```
stats <- stargazer(as.data.frame(Base), type = "text",
  out = "tabla1.tex",
  title = "Tabla 1 - Estadística Descriptiva",
  digits = 1,
  covariate.labels = c("NOR-Hombres", "NOR-Mujeres",
    "US-Mujeres",
    "US-Hombres", "Corrupción", "IDH", "Desempleo"),
  flip = TRUE)

##
## Tabla 1 - Estadística Descriptiva
## =====
## Statistic NOR-Hombres NOR-Mujeres US-Mujeres US-Hombres Corrupción IDH Desempleo
## -----
## N          196          196          196          196          165          156          70
## Mean       370.0       437.9       35,991.9       34,153.9       4.0          0.7          7.1
## St. Dev.   1,002.9     994.1       89,276.3       85,572.5       2.1          0.2          5.4
## Min        0          0          0          0          1.1          0.3          0.4
## Pctl(25)    6          7.5       205.8       235.8       2.4          0.5          3.4
## Pctl(75)   214.5       266       26,493.5       28,110.5       5.1          0.8          9.8
## Max        9,940      6,785       672,968       675,837       9.3          0.9         26.8
## -----
```

En primer lugar, cabe decir que el número de observaciones para las variables dependientes de este estudio (migración de hombres y mujeres para Noruega y Estados Unidos) es 196. En cuanto a las variables independientes, es decir, corrupción, IDH y desempleo, no se tiene el mismo número de observaciones. La variable que tiene más valores faltantes es la relacionada con el desempleo (con 126 valores), seguida a esta, se encuentra el índice de desarrollo humano con 40 valores faltantes y por último, la variable proxy de corrupción, que tiene 31 valores faltantes.

Ahora, teniendo en cuenta la media, se puede decir que Estados Unidos tiene un flujo más grande de migración calificada que Noruega, representando aproximadamente el 1% de la estadounidense. Además, se evidencia, para los dos países, que la migración altamente cualificada femenina es mayor que la masculina. Por otro lado, el índice de percepción de corrupción muestra que en un escala de 0-10, que es la utilizada para el periodo de estudio (2010), captura una media global de 4. Sin embargo, para Estados Unidos se tiene un índice de corrupción de 7.1 y para Noruega de 8.6.

Por su parte, el índice de desarrollo humano está medido en una escala de 0-1. Para este caso, Estados Unidos tiene un índice más alto (0.916) que Noruega (0.940) y la media global es de 0.7. Ambos de los países en estudio están catalogados como países con alto desarrollo Humano, al tener un valor mayor a 0.8. Por último, no se conoce ni para Estados Unidos, ni para Noruega, la tasa de desempleo en el año 2010, en otras palabras, hacen parte de los valores faltantes.

En cuanto a la desviación estándar, ésta nos indica que el número de migrantes altamente cualificados, tanto para hombres como mujeres en Noruega, se desvía de la media aproximadamente 1000 personas. En Estados Unidos, la desviación estándar es mucho más amplia, con un promedio de 87000 personas.

Revisando esta estadística para las variables independientes, primero, se encuentra que el puntaje para el índice de corrupción se desvía de la media en 2.1 puntos; segundo, el puntaje para el índice desarrollo humano se desvía en 0.2 puntos en su escala y tercero la tasa de desempleo se desvía de su media en 5.4%.

El valor mínimo para las variables dependientes es de 0, es decir, ninguna persona migró a Estados Unidos o Noruega desde otro país; el valor máximo para la migración hacia Noruega es mayor para los hombres

(9940 de Polonia) que para las mujeres (6785 de Suecia). En Estados Unidos el máximo también es mayor para los hombres (675837 de México) que para las mujeres (672968 de Filipinas). No obstante, como se dijo anteriormente se tiene una media mayor para las mujeres que para los hombres.

La variable proxy de corrupción tiene un valor mínimo de 1.1 en Somalia y un valor máximo de 9.3 en Dinamarca, Singapur y Nueva Zelanda. Por otro lado, el índice de desarrollo humano muestra un valor mínimo de 0.3 en Nigeria y uno máximo de 0.9 en Suiza y para la tasa de desempleo se tiene un mínimo de 0.4 en Qatar y un máximo de 26.8 en Lesoto.

Como última estadística descriptiva se tiene a los percentiles 25 y 75, también llamados cuartiles 1 y 3, respectivamente. El primero de estos muestra que el 25% de los datos a) de migración hacia Noruega es menor o igual aproximadamente a 6.7 personas, b) de migración hacia Estados Unidos es menor o igual a aproximadamente 221 personas, c) del índice de percepción de corrupción es menor o igual a 2.4, d) del índice de desarrollo humano es menor o igual a 0.5 y e) de la tasa de desempleo es menor o igual a 3.4. Por su parte, el cuartil 3 muestra que el 75% de los datos son iguales o menores a) en la migración hacia Noruega a aproximadamente 240 personas, b) en la migración hacia Estados Unidos a aproximadamente 27301 personas, c) en el índice de corrupción a 5.1 puntos, d) en el índice de desarrollo humano a 0.8 y e) en la tasa de desempleo a 9.8.

Punto 6

Se realiza un histograma para la variable dependiente y para la principal variable independiente. Las gráficas se realizan usando el paquete ggplot.

Variables dependientes

Debido a la necesidad de realizar distintos modelos con 4 variables dependientes:

Y_1, Y_2, Y_3, Y_4

Bins optimo

$1 + (\log(196)/\log(2)) = 9$

La nomenclatura usada para expresar la cantidad de personas altamente cualificadas por país de destino y por genero será *No.Cualificados*. Adicionalmente, para reflejar el número de nacionalidades o de país de origen se usara el término *Países*.

```
a6 <- ggplot(data= Base, mapping = aes(x=HighMenNor))+
  geom_histogram(aes(HighMenNor), bins = 9, fill = "darkgoldenrod1", color = "black")+
  labs(title="NOR-Hombres", x= "No.Cualificados", y= "Países")+
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title.x = element_text(family = "serif", size = 16), axis.title.y = element_text(family = "serif", size = 16))
```

```
b6 <- ggplot(data= Base, mapping = aes(
  x=HighWomNor,
))+
  geom_histogram(aes(HighWomNor), bins = 9, fill = "darkolivegreen1", color = "black")+
  labs(title="NOR-Mujeres", x= "No.Cualificados", y= "Países")+
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title.x = element_text(family = "serif", size = 16), axis.title.y = element_text(family = "serif", size = 16))
```

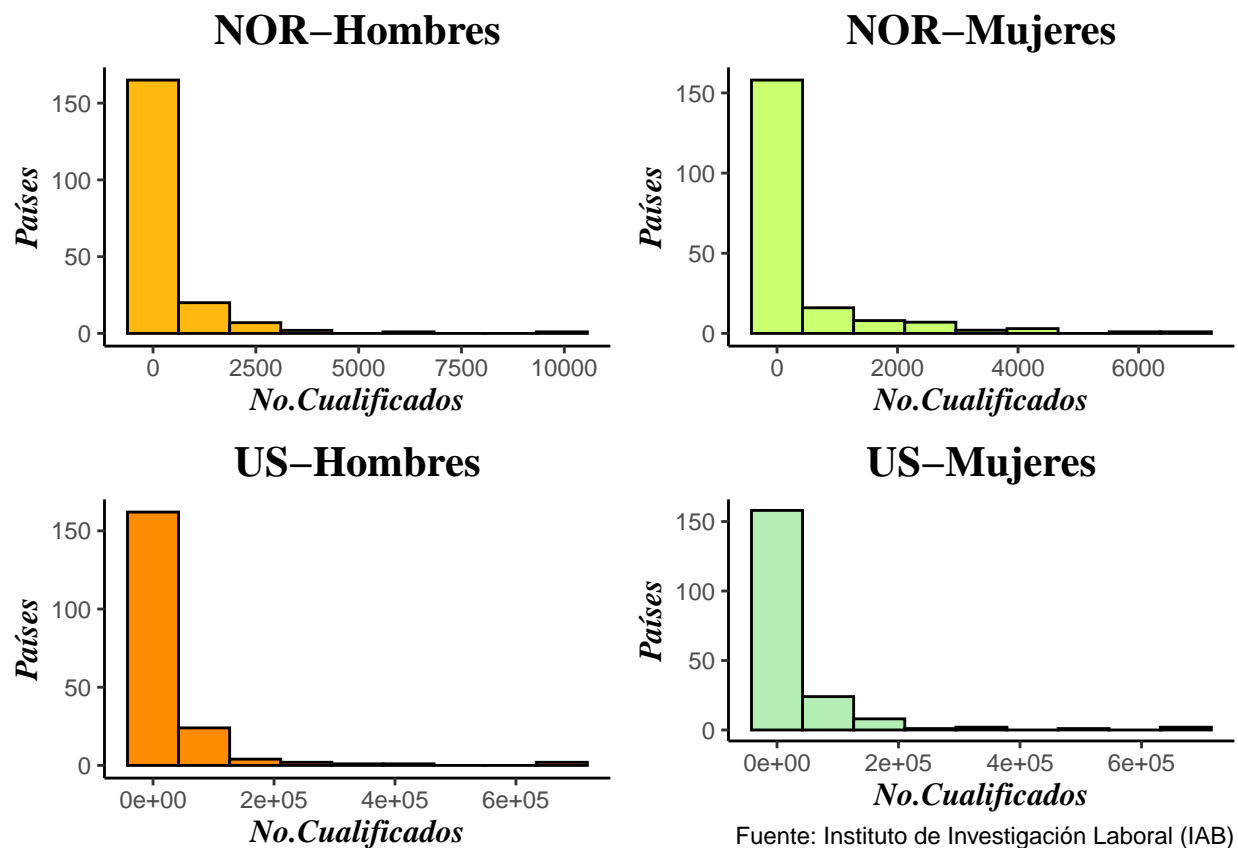


```
c6 <- ggplot(data= Base, mapping = aes(
  x=HighMenUS,
))+
geom_histogram(aes(HighMenUS), bins = 9, fill = "darkorange", color = "black")+
labs(title="US-Hombres", x= "No.Cualificados", y= "Países")+
theme_classic()+
theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title..

''r
d6 <- ggplot(data= Base, mapping = aes(
  x=HighWomUS))+
geom_histogram(aes(HighWomUS), bins = 9, fill = "darkseagreen2", color = "black" )+
labs(title="US-Mujeres", x= "No.Cualificados", y= "Países" , caption = "Fuente: Instituto de Investig
theme_classic()+
theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title..
```

Uniendo las distintas gráficas obtenemos:

```
ggarrange(a6, b6, c6, d6)
```



Para los 2 países y los 2 géneros evidenciamos una relación similar, por no decir igual. Donde no hay una discriminación por sexo en cuanto a distribución de los datos, tanto para Noruega como para los Estados Unidos.

Y algo quizás más relevante es que hay una concentración grande de los datos en los 4 escenarios (Hombre-Noruega, Mujer-Noruega, Hombre-Estados Unidos y Mujer-Estados Unidos) de más de 150 países de origen

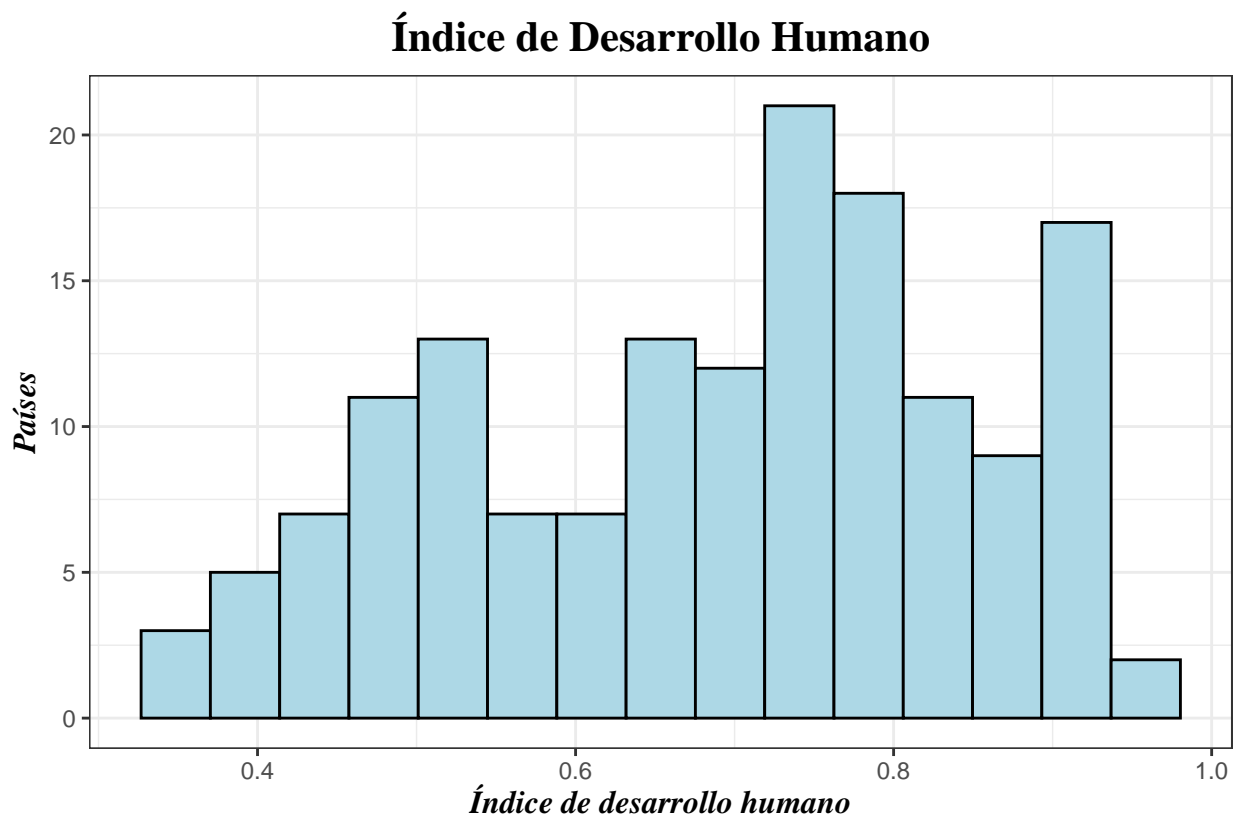
o nacionalidad en la cual hubo 0 personas altamente calificadas que inmigraron a estos dos países o fueron pocas.

Se encuentran una excepciones que se visualizan casi como lineas en los 4 gráficos.

Variable Independiente

```
ggplot(data= Base, mapping = aes(x=Development))+  
  geom_histogram(aes(Development), bins = 15, fill= "lightblue", color = "black")+  
  labs(title="Índice de Desarrollo Humano", x= "Índice de desarrollo humano", y= "Países" , caption = "I")  
  theme_bw()+  
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title..
```

```
## Warning: Removed 40 rows containing non-finite values (stat_bin).
```



Fuente: UNESCO Institute for Statistics

Para el caso del índice de Desarrollo Humano se encuentra una distribución “trimodal”, aunque tiende un poco más a una bimodal con un grupo exceptuado.

Esto remarca 3 bloques: países de ingresos bajos que ocupan la primera campana, países de ingreso medio o economías emergentes que ocupan la segunda campana, que son la mayoría, y por último un grupo exceptuado de éstas dos campanas o bien una tercera campana empinada referente a economías avanzadas (ingresos altos) con altos índices de desarrollo humano.

Punto 7

Para las variables cuantitativas (continuas y discretas), se calcula la correlación entre cada pareja de variables y se presentan estas correlaciones en una matriz de correlaciones.

```
PNor <- Base %>% select(c("HighMenNor", "HighWomNor", "Corruption", "Development", "Unemployment"))
ggpairs(PNor)+
  theme_grey()+
  theme(strip.text = element_text(family = "serif", face = "bold.italic", size = 9), strip.background

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 31 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 40 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 126 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 31 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 40 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 126 rows containing missing values

## Warning: Removed 31 rows containing missing values (geom_point).

## Warning: Removed 31 rows containing missing values (geom_point).

## Warning: Removed 31 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 46 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 128 rows containing missing values

## Warning: Removed 40 rows containing missing values (geom_point).

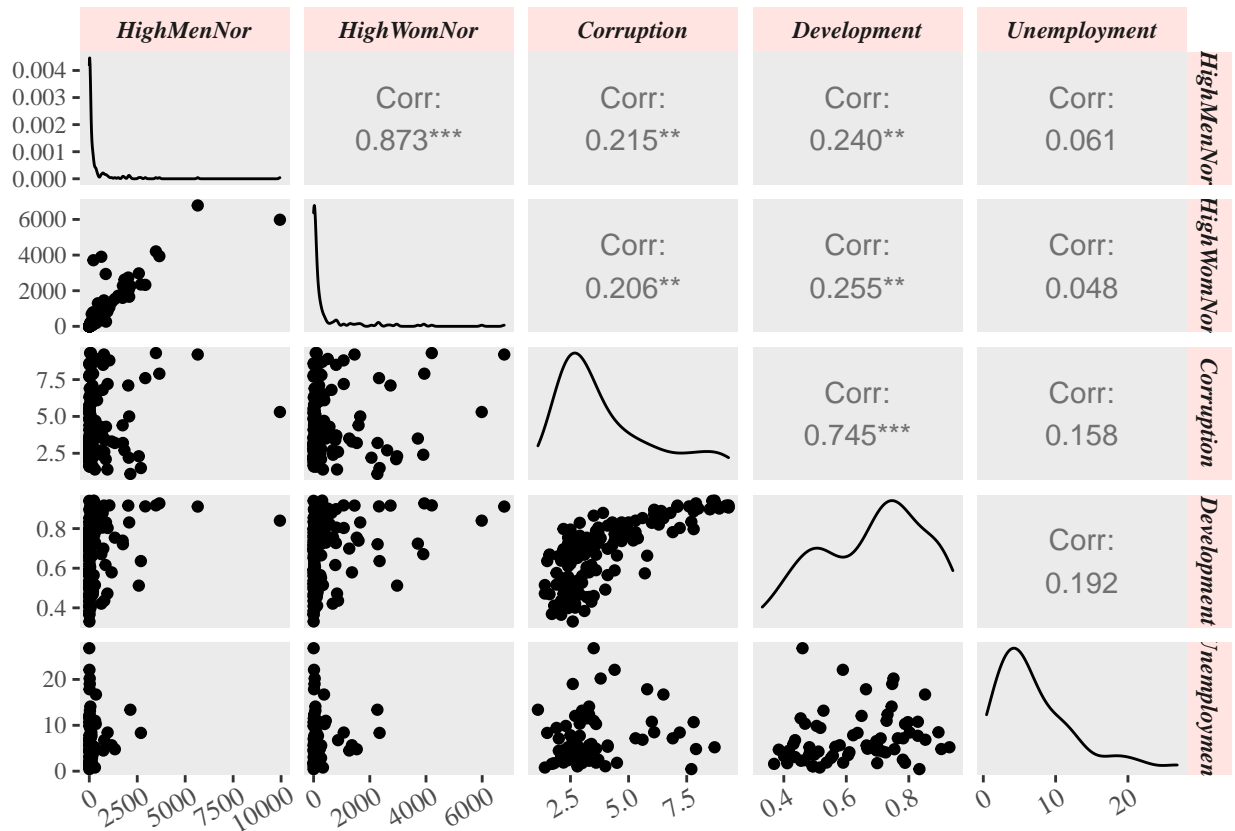
## Warning: Removed 40 rows containing missing values (geom_point).

## Warning: Removed 46 rows containing missing values (geom_point).

## Warning: Removed 40 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 129 rows containing missing values
```

```
## Warning: Removed 126 rows containing missing values (geom_point).
## Warning: Removed 126 rows containing missing values (geom_point).
## Warning: Removed 128 rows containing missing values (geom_point).
## Warning: Removed 129 rows containing missing values (geom_point).
## Warning: Removed 126 rows containing non-finite values (stat_density).
```



Lo que salta a primera vista es la fuerte correlación que se presenta entre *HigMenNor* y *HighWomNor* que indican la cantidad de inmigrantes masculinos y femeninos respectivamente, que son altamente cualificados y viajaron a Noruega en el año 2010 con una correlación de 0.87, muy cercana a 1.

Con ello podemos inferir que las brechas en cuanto a fuga de cerebros no es sustancial, pues están dadas por unas condiciones iniciales de los países tanto de origen como el de destino, que en este caso es el de Noruega.

Otro dato que es inferible inclusive sin los datos es el estándar de corrupción medido en el Índice de Percepción de Corrupción y el nivel de desarrollo medido en el Índice de Desarrollo Humano. El primero con una escala de 0 a 10, donde 10 se percibe como el menos corrupto y con mayor transparencia, mostrando que tener un buen puntaje está correlacionado de igual manera con tener un IDH alto.

Consecuentemente hay una correlación positiva pero débil de corrupción (mejor puntaje). Tanto hombres como mujeres prefieren inmigrar a países como Noruega relacionandose quizás con el deseo de desarrollarse académicamente y desenvolver su carrera laboral en el exterior.

Por último, el desempleo tiene una correlación casi nula del 0.06 y el 0.048 con la inmigración tanto de hombres como de mujeres a Noruega.

```

PUS <- Base %>% select(c("HighMenUS", "HighWomUS", "Corruption", "Development", "Unemployment"))
ggpairs(PUS)+
  theme_classic()+
  theme(strip.text = element_text(family = "serif", face = "bold.italic", size = 9), strip.background =

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 31 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 40 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 126 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 31 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 40 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 126 rows containing missing values

## Warning: Removed 31 rows containing missing values (geom_point).

## Warning: Removed 31 rows containing missing values (geom_point).

## Warning: Removed 31 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 46 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 128 rows containing missing values

## Warning: Removed 40 rows containing missing values (geom_point).

## Warning: Removed 40 rows containing missing values (geom_point).

## Warning: Removed 46 rows containing missing values (geom_point).

## Warning: Removed 40 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 129 rows containing missing values

## Warning: Removed 126 rows containing missing values (geom_point).

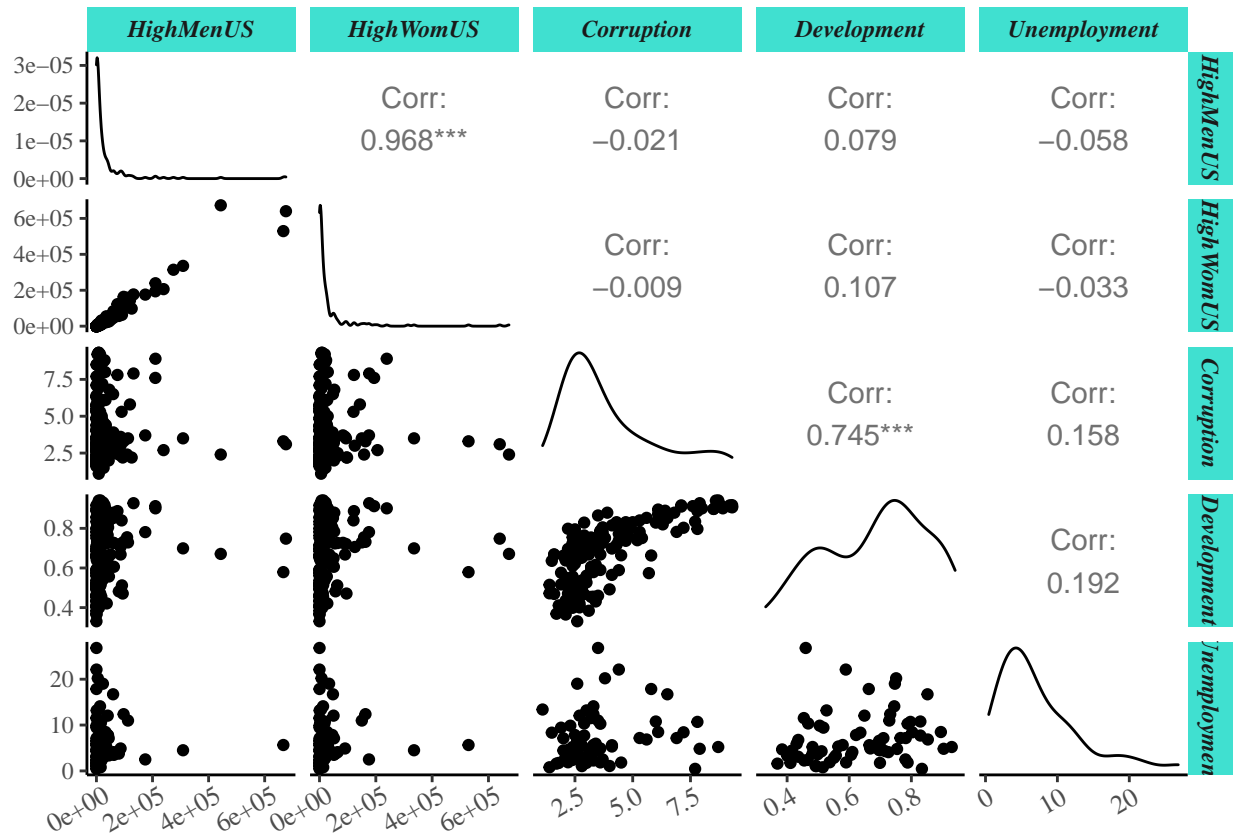
## Warning: Removed 126 rows containing missing values (geom_point).

```

```
## Warning: Removed 128 rows containing missing values (geom_point).

## Warning: Removed 129 rows containing missing values (geom_point).

## Warning: Removed 126 rows containing non-finite values (stat_density).
```



A diferencia con Noruega como destino de mano de obra altamente cualificada, Estados Unidos reduce la brecha mínima que existía en el caso anteriormente detallado, donde la correlación era de 0.87 y en este caso llega hasta un 0.96. No obstante, hasta este punto no se puede asegurar si la brecha favorezca algún género.

No obstante, de manera curiosa por parte del índice de Percepción de Corrupción se muestra una correlación si bien insignificante, negativa de -0.021 para hombres y de -0.009. Sucede lo mismo con la variable del desempleo donde antes era positiva y con una correlación muy pequeña, en este caso se vuelve negativa pero sigue siendo pequeña de -0.05 y -0.03 respectivamente.

Punto 8

Se realiza una gráfica que permite relacionar la variable dependiente y la principal variable independiente de interés. Como se tienen varias variables independientes de interés, se realiza una gráfica para mostrar la relación entre cada una de estas variables y la variable dependiente. Además, se realicen las transformaciones en las variables que consideren necesarias para mostrar de mejor manera la información en estas gráficas. Las gráficas se realizan usando el paquete *ggplot* y se siguen las siguientes recomendaciones:

- Si las dos variables son cuantitativas el gráfico debe ser un diagrama de dispersión.

- Si las dos variables son categóricas el gráfico debe ser un gráfico de barras que permita visualizar las dos variables.
- Si una variable es cuantitativa y la otra categórica, en un mismo gráfico realice un polígono de frecuencia, una densidad o un diagrama de cajas de la variable cuantitativa para cada una de las categorías de la variable categórica. Por ejemplo, si fueran a analizar la relación entre el tipo de colegio y puntaje, debería hacer un polígono de frecuencias, una densidad o un diagrama de cajas para el puntaje para los colegios públicos y otro para los colegios privados. Ambos polígonos o densidades deberían estar en la misma gráfica (i.e. mismo plano cartesiano) para poder apreciar las diferencias en la distribución de los puntajes de los estudiantes de ambos tipos de colegios. Si la variable categórica tuviera más de tres categorías, la gráfica podría hacerse con una función *facet_*.

Poseemos 4 variables dependientes seccionadas en Noruega y Estados Unidos las cuales también se encuentran divididas por si los inmigrantes fueron mujeres o hombres. La principal variable de interés es el índice de desarrollo humano.

```
a8 <- ggplot(data = Base,
             mapping = aes(x = Development, y = HighMenUS)) +
  geom_point(color = "coral", shape = 19, alpha = 0.7, size = 4) +
  scale_y_continuous(labels = label_number()) +
  labs(title = "Dispersograma US-Hombres", x = "Índice de Desarrollo Humano", y = "No. Cualificados")+
  theme_pander()+
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title.x = element_text(family = "serif", size = 16), axis.title.y = element_text(family = "serif", size = 16))
```

```
b8 <- ggplot(data = Base,
             mapping = aes(x = Development, y = HighWomUS)) +
  geom_point(color = "cornflowerblue", shape = 19, alpha = 0.7, size = 4) +
  labs(title = "Dispersograma US-Mujeres", x = "Índice de Desarrollo Humano", y = "No. Cualificados")+
  scale_y_continuous(labels = label_number())+
  theme_pander()+
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title.x = element_text(family = "serif", size = 16), axis.title.y = element_text(family = "serif", size = 16))
```

```
c8 <- ggplot(data = Base,
             mapping = aes(x = Development, y = HighWomNor)) +
  geom_point(color = "olivedrab1", shape = 19, alpha = 0.7, size = 4) +
  scale_y_continuous(labels = label_number()) +
  labs(title = "Dispersograma NOR-Mujeres", x = "Índice de Desarrollo Humano", y = "No. Cualificados")+
  theme_pander()+
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title.x = element_text(family = "serif", size = 16), axis.title.y = element_text(family = "serif", size = 16))
```

```
d8 <- ggplot(data = Base,
             mapping = aes(x = Development, y = HighMenNor)) +
  geom_point(color = "plum2", shape = 19, alpha = 0.7, size = 4) +
  scale_y_continuous(labels = label_number()) +
  labs(title = "Dispersograma NOR-Hombres", x = "Índice de Desarrollo Humano", y = "No. Cualificados")+
  theme_pander()+
  theme(plot.title = element_text(hjust = 0.5, face = "bold", family = "serif", size = 16), axis.title.x = element_text(family = "serif", size = 16), axis.title.y = element_text(family = "serif", size = 16))
```

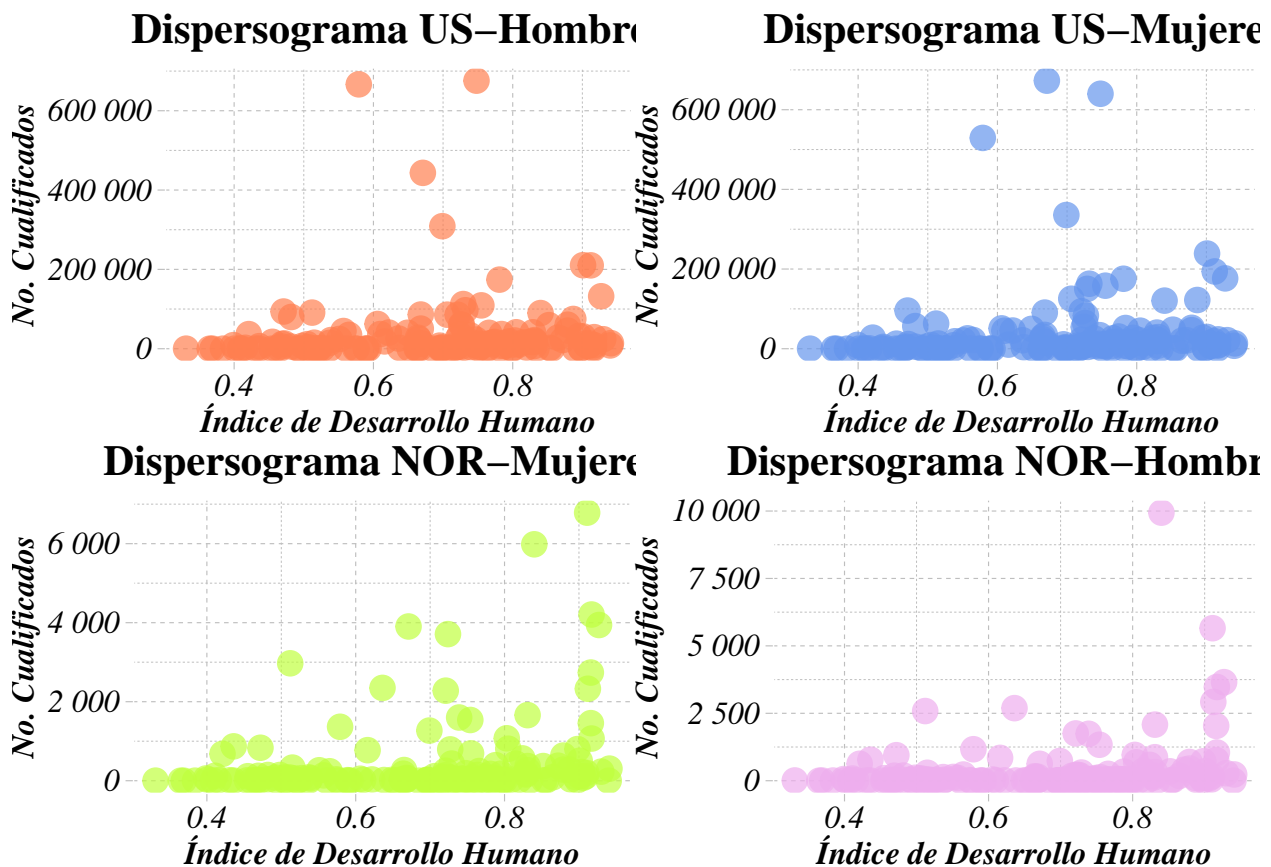
```
ggarrange(a8,b8,c8,d8)
```

```
## Warning: Removed 40 rows containing missing values (geom_point).
```

```
## Warning: Removed 40 rows containing missing values (geom_point).
```

```
## Warning: Removed 40 rows containing missing values (geom_point).
```

```
## Warning: Removed 40 rows containing missing values (geom_point).
```



Visualmente encontramos respuesta a la pregunta que se termino planteando en el punto anterior sobre la brecha que existia entre hombres y mujeres altamente cualificados para migrar a Noruega y a Estados Unidos, donde la primera contaba con un 0,87 y la segunda con un 0,96.

En el caso de Estados Unidos el dispersograma muestra una convergencia del comportamiento de inmigración masculina y femenina hacia este país. Sin embargo en el caso Noruega, las mujeres altamente cualificadas tienen un mayor flujo migratorio dado cualquier nivel de desarrollo del país de origen. La única excepción a la regla es una nacionalidad de hombres que para el 2010 se acercó a 10.000 inmigrantes masculinos a Noruega.

Esto demuestra la razón inicial de poner a Noruega en comparación con Estados Unidos y es que su selección se basó en que está posicionado como un referente en eliminación de brechas de género y uno de los mejores países para vivir siendo mujer tanto en la revista Britanica The Economist como el Gender Index de National Geographic.

Punto 9

Usando la información presentada en las preguntas 6 a 9, ¿qué se concluye sobre la distribución de las variables de interés y sobre la relación entre estas variables?

Se evidencian las grandes brechas de desarrollo entre países y oportunidades de acceso para convertirse en mano de obra altamente cualificada. Primeramente, se muestra que para el análisis de Estados Unidos y

Noruega como países receptores de esta mano de obra altamente cualificada y distinguiendo por género, hay una gran concentración de pocos inmigrantes o directamente la carencia de estos hacia los dos países analizados.

Conjuntamente se evidencia esa triple división en desarrollo humano que puede imposibilitar lo relatado en el párrafo anterior, ya que los países de ingresos bajos y medios son mayoría.

Un segundo detalle para evidenciar se da en las correlaciones de las variables, el caso de que no hay una discriminación por sexo independiente del destino, pero si existe una brecha en la cual dadas unas condiciones iniciales las mujeres se ven más incentivadas a inmigrar como mano de obra altamente cualificada.

Destacando en Noruega que, independientemente de los niveles de desarrollo de país de origen, inmigran más a Noruega que los hombres cosa que se había mencionado siendo este país uno de los mejores poseedores tanto en un top de los mejores países para ser mujer tanto de The Economist como de National Geographic.