

# Diplomado de Análisis Estadístico usando R

## Módulo 3: Exploración de datos Guía de Ejercicios

**Profesor:** *Víctor Macías E.*

Esta guía está constituida por dos partes:

- Sección I explora los datos de población, expectativa de vida al nacer, PIB per cápita y coeficiente de Gini obtenidos de “World Development Indicators” (WDI) en la Guía de Ejercicios 2.
- Sección II presenta una serie de ejercicios relacionados a manipulación de fechas, factores y strings.

### Sección I

#### Pregunta 1

Nombre a las variables *SP.POP.TOTL*, *SP.DYN.LE00.IN*, *NY.GDP.PCAP.PP.CD* y *SI.POV.GINI* como *population*, *lifeExpectancy*, *gdppc* y *gini*, respectivamente.

#### Pregunta 2

Ordene los datos en orden ascendente primero por *country* y luego por *year*

#### Pregunta 3

Cambie los nombres de las regiones a español. Por ejemplo, cambie la región “Middle East & North Africa” a “Medio Oriente y Norte de Africa”.

#### Pregunta 4

¿Cuántos países tiene la base de datos? ¿Cuántas observaciones por región? ¿Cuántos países por región?

#### Pregunta 5

Identifique los 5 países con mayor población el año 2019

## Pregunta 6

Identifique los 10 países con mayor crecimiento de su población entre los años 2018 y 2019.

## Pregunta 7

Identifique los 10 países con menor PIB per cápita el año 2019

## Pregunta 8

Para cada una de las regiones del mundo, presente una tabla con estadística descriptiva para las variables PIB per cápita y expectativa de vida para el año 2018.

## Pregunta 9

Construya una tabla con el número de *NAs* para las variables *population*, *gdppc*, *lifeExpectancy* y *gini* en cada uno de los años incluidos en los datos.

## Pregunta 10

Para cada una de las regiones del mundo, determine los valores mínimos y máximos del PIB per cápita el año 2018. Además, identifique los países con los mínimos y máximos del PIB per cápita en cada una de las regiones.

## Pregunta 11

Compare las expectativas de vida de El Salvador, Guatemala, Honduras y México en 1980 y 2018

## Pregunta 12

Guarde los datos como *wbData3.rds*

## Sección II

### Pregunta 1

A partir de la información contenida en la siguiente tabla, genere un *date* y *datetime*

year	mes	dia	hora	minuto	segundo
2018	5	20	19	23	44
2019	8	12	4	45	12
2020	6	4	6	5	53

### Pregunta 2

Extraer el año, mes, día, número de día y día de la semana de las siguientes fechas: *2018-05-20 12:56:32*, *2019-08-12 04:23:45* y *2020-06-04 19:43:09*.

### Pregunta 3

¿Cuántos días hay entre el 10 de enero del 2012 y el día de hoy?

### Pregunta 4

La variable *mes* tiene los siguientes valores: *Marzo*, *Abril*, *Febrero*, *Mayo*, *Enero*, *Junio*. Defínala como factor manteniendo el orden cronológico de los meses.

### Pregunta 5

El vector *x* contiene una lista de países, los cuales comienzan en mayúsculas y minúsculas y, además, incluyen caracteres que no son letras.

```
x <- c("argentina(1)", "Brazil", "el salvador", "guatemala", "Honduras", "Uganda")
```

- (a) Elimine cualquier caracter que no sea una letra
- (b) Cambia a mayúscula la primera letra de cada uno de los países cuando sea necesario
- (c) Cambiar a mayúsculas todas las letras de cada uno de los países
- (d) ¿Cuántos nombres de países comienzan con una vocal? ¿Qué países comienzan con una vocal?
- (e) ¿Cuántos nombres de países comienzan con una consonante? ¿Qué países comienzan con una consonante?
- (f) Liste los países cuyos nombres terminan en *a*
- (g) Liste los países cuyos nombres comienzan en *A* o *U*
- (h) ¿Cuántas letras tiene cada país?
- (i) ¿Cuántas vocales, en promedio, tiene el nombre de cada país?

- (j) Obtenga la letra del medio de cada uno de los países ¿Qué recomienda hacer si el largo del string es un número par?
- (k) Liste los países que comienzan con 2 consonantes
- (l) Liste los países cuyos nombres comienzan y terminan con una vocal.