



**Profesores:**

Dr. Alberto Prado

Dr. Ulises Olivares Pinto

# CURSO INTER-SEMESTRAL INTRODUCCIÓN A



**PUBLICO:** PROFESORES,  
INVESTIGADORES Y ESTUDIANTES

**REQUISITOS:** NINGUNO

**DURACIÓN DEL CURSO:** 20 HRS

**FECHAS:** 17-21 DE AGOSTO 2020

# INTRODUCCIÓN

4 HORAS

- ¿Qué es R?
- ¿Cómo funciona?
- Objetos R
- R Studio
- Funciones básicas

# ¿Qué es R?



- R es un software gratuito para el análisis y visualización de datos
- R es un lenguaje de programación
- R es un ambiente de trabajo
  
- R corre en Windows, Mac y LINUX
- R es distribuido y desarrollado por un grupo mundial de matemáticos llamado el *R Development Core Team* con sede en el Instituto de Estadística y Matemáticas de la Universidad de Vienna, Austria.

Ihaka R.& Gentleman R. 1996. R :a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5: 299-314.

## Apuntador

**>** R espera tus comandos

**+** R espera que termines tu comando

## Comandos básicos

**<-** Asignación

**+** suma

**-** resta

**\*** multiplicación

**/** división

**^** exponente

**()** paréntesis

**log()** logaritmo natural

**sqrt()** raíz cuadrada

## Apuntador

**>** R espera tus comandos

**+** R espera que termines tu comando

## Operadores aritméticos básicos

**<-** Asignación

**+** suma

**-** resta

**\*** multiplicación

**/** división

**^** exponenent

**()** paréntesis

**log()** logaritmo natural

**sqrt()** raíz cuadrada

## Ejercicio:

**a <- 2**

**b <- 3**

**c <- a+b**

¿dónde esta c?

## Apuntador

> R espera tus comandos

+ R espera que termines tu comando

## Comandos básicos

<- Asignación

+ suma

- resta

\* multiplicación

/ división

^ exponenent

() paréntesis

log() logaritmo natural

sqrt() raíz cuadrada

## Ejercicio:

a <- 2

b <- 3

c <- a+b

¿dónde esta c?

Necesitamos llamar a **c**

## ¿Cómo funciona R?

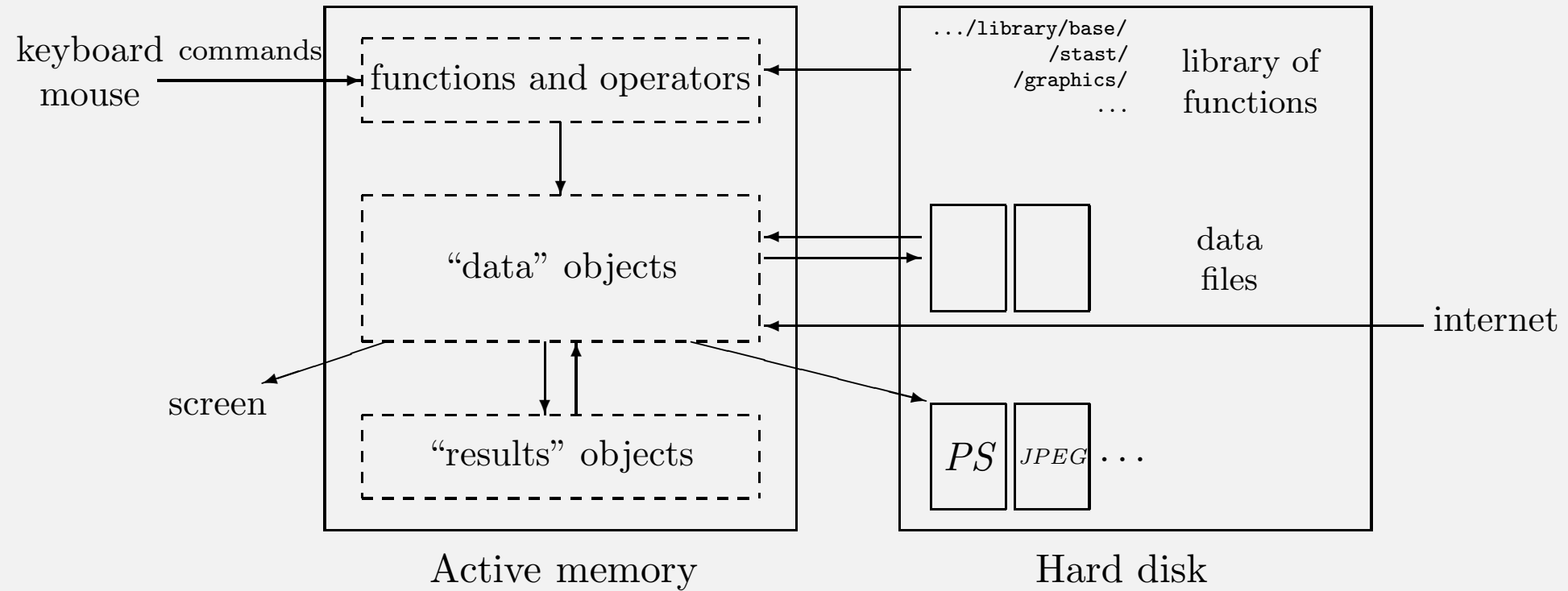


Figure 1: A schematic view of how R works.

## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función



## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Vector:** Es la estructura mas sencilla de datos.

Una colección de datos del mismo tipo.

Son definidos por la función **c()**

```
X <- c(1,2,3,4,5)
```

```
Y <- c("Hugo", "Paco", "Luis")
```

```
X[3]
```

```
Y[2]
```

## Tipos de datos

Tipo	Abreviaciones	Ejemplo
Numérico	num	1.12
Entero	int / double	2
Texto	chr	“Hugo”
Factor	fac	“Rojo”
Fecha	date	“2020-01-20”
Lógico	logical	TRUE

## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Matriz:** Es una colección de datos del mismo tipo en un número específico de filas y columnas.

Indexación: nombre de la matriz[número de fila , número de columna]

```
mat <- matrix(1:9, nrow = 3, ncol = 3)
```

```
mat[2,1]
```

## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Data frame:** Es una estructura bi-dimensional de datos, dónde las filas son las observaciones y las columnas son las variables.

A diferencia con la matriz, las variables pueden ser de diferentes tipos de datos.

```
dat <- data.frame(id = letters[1:10], x = 1:10, y = 11:20)  
dat
```

**str()** función para revisar la estructura interna de un objeto R

## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Lista:** Es un contenedor de cualquier tipo de objeto R, inclusive otras listas. Se puede entender como un vector de cualquier tipo de dato u otro objeto R.

```
x <- list(1, "a", TRUE, 2+4*i)  
x
```

**rm(list=ls())** comando para borrar todas las listas del ambiente R

## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Función:** Es una secuencia de líneas para realizar una acción específica. R ya tiene una vasta cantidad de funciones incluidas, pero se pueden crear nuevas. El nombre de la función va seguido de paréntesis.

`nombreFunción(Objeto,Argumentos)`

Ejemplos de funciones existentes :

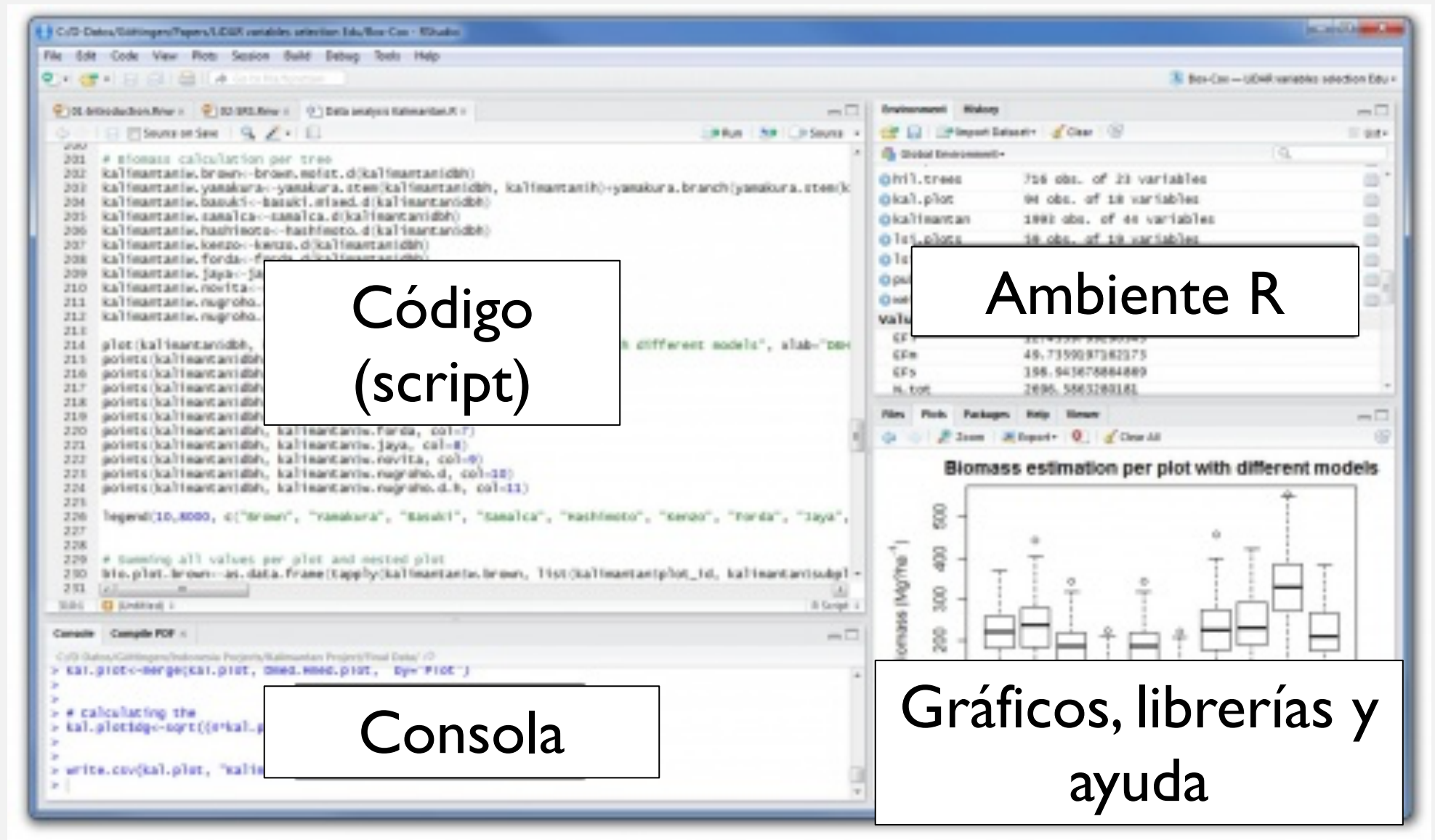
`mean()`

`sd()`

`length()`

# R Studio

R Studio es un software que facilita la programación en R



## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Función:** Es una secuencia de líneas para realizar una acción específica. R ya tiene una vasta cantidad de funciones incluidas, pero se pueden crear nuevas.

Ejemplos de como crear una función:

```
fahrenheit_a_celsius <- function(temp_F) {  
  temp_C <- (temp_F - 32) * 5 / 9  
  return(temp_C) }
```

Ayuda: ?nombre de la función



## Algunos objetos R

- Vector
- Matriz
- Data frame
  - Variable
- Lista
- Gráfico
- Función

**Función:** Es una secuencia de líneas para realizar una acción específica. R ya tiene una vasta cantidad de funciones incluidas, pero se pueden crear nuevas.

Ejemplos de como crear una función:

```
fahrenheit_a_celsius <- function(temp_F) {  
  temp_C <- (temp_F - 32) * 5 / 9  
  return(temp_C) }
```

### Ejercicios:

Haz tres funciones

- 1) Recibe 2 números y calcula la multiplicación de las raíces cuadradas
- 2) Saque el promedio de un vector
- 3) Sume 1 a los valores de un vector y después calcule el logaritmo natural

## Estructura sugerida para el encabezado de un script

```
rm(list=ls())          #### Limpiar memoria
setwd("~/Desktop/curso R 2020") #### Directorio de trabajo
#####
#### Autor y Fecha
#### Versión de R con la que se esta trabajando
#### Curso Intersemestral ENES J UNAM
#### Detalles de lo que se va hacer
```

## Importar datos

`read.csv(“nombreArchivo.csv”)`

`read.csv2(“nombreArchivo.csv”)`

`read.table(“nombreArchivo.txt”)`



## Operadores relacionales

Operador	Definición
<	Menor que
>	Mayor que
<=	Menor o igual a
>=	Mayor o igual a
==	Igual a
!=	No igual a

## Operadores lógicos

Operador	Descripción
!	Logico No (negación)
&	AND / Y
	OR / O

## Operadores varios

Operador	Descripción
\$	Extraer elemento de una lista
~	En función de

## Usando Indexación con operadores

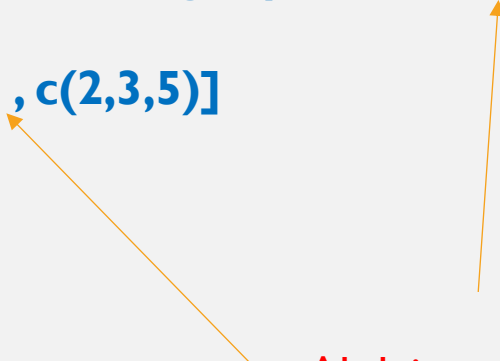
**Nombre.data.frame[filas,columnas]**

**iris[iris\$Especie=="setosa" , ]**

**iris[iris\$Long.sepalo<=4.5 , ]**

**Iris[ , c(2,3,5)]**

¡Al dejar en blanco  
se escogen todas las  
filas o columnas!

Two orange arrows originate from the text. One arrow points from the text to the blank space before the comma in 'Iris[ , c(2,3,5)]'. The other arrow points from the text to the blank space after the comma in 'iris[iris\$Long.sepalo<=4.5 , ]'.

## Ejercicio:

##### Importar datos

```
read.csv("iris.csv")
```

```
iris <- read.csv("iris.csv")
```

## Revisar estructura del Data Frame

```
str(iris)
```

## Eliminar columna

```
iris <- iris[,-1]
```



## Ejercicio:

## cambiar nombres de las columnas

**names(iris)** # indica los nombres

**names(iris) <- c("Long.sepalo","Ancho.sepalo","Long.petal","Ancho.petal","Especie")**

## Nombre o numero de fila

**rownames(iris)**

## Calcular Media, mediana y desviación estándar

**mean(iris\$Long.petal)**

**median(iris\$Ancho.sepalo)**

**sd(iris\$Ancho.petal)**

#### numero de elementos por categoría

**table(iris\$Especie)**

## Ejercicio:

## gráficos descriptivos básicas

**hist(iris\$Long.sepalo)**

**plot(iris)**

**dotchart(iris\$Long.sepalo, groups=iris\$Especie)**

**plot(x=iris\$Long.sepalo, y=iris\$Ancho.sepalo)**

**plot(x=iris\$Long.sepalo, y=iris\$Ancho.sepalo, col=iris\$Especie)**

**plot(x=iris\$Long.sepalo, y=iris\$Ancho.sepalo, col=iris\$Especie, main= "Mi gráfica", xlab="Longitud del sepalo (cm)", ylab="Ancho del sepalo (cm)")**

# caja y bigote

**boxplot(iris\$Long.sepalo~iris\$Especie)**

## Ejercicios en casa:

1) Usa el data frame iris para hacer el siguiente ejercicio.

Excluye todas las líneas con el ancho del sépalo menor a 2.5 cm.

Haz una gráfica de dispersión de la longitud del pétalo en función de la longitud del sépalo y colorea los puntos por especie.

Cambia los títulos de los ejes.

2) Usa el data frame Wage para hacer el siguiente ejercicio.

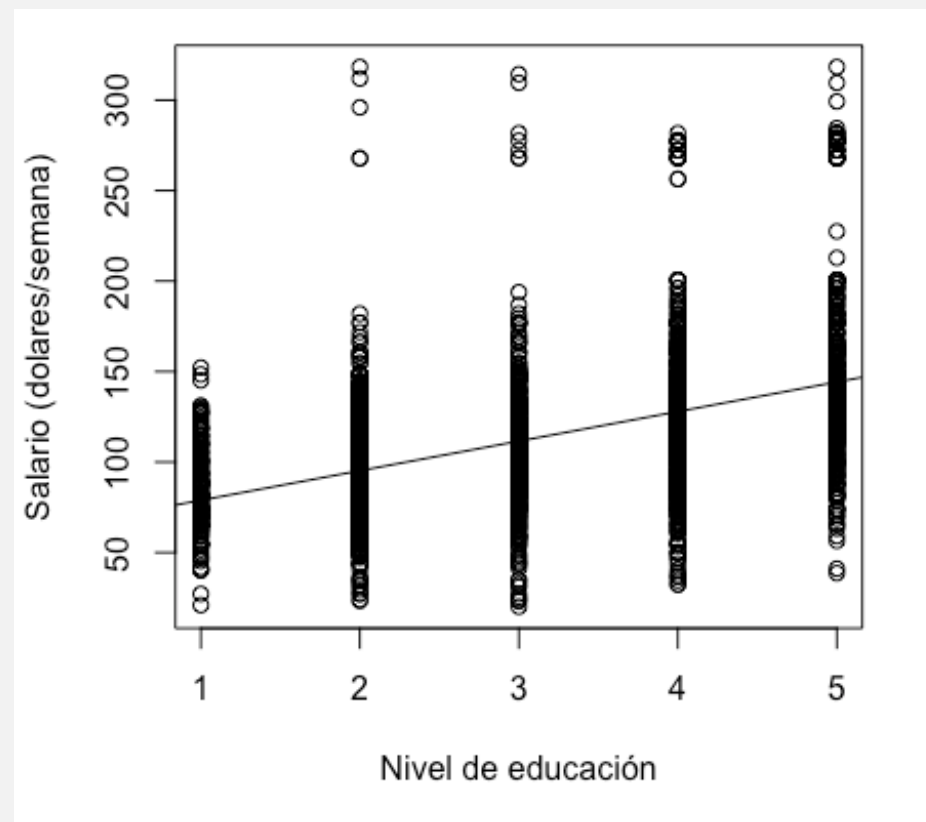
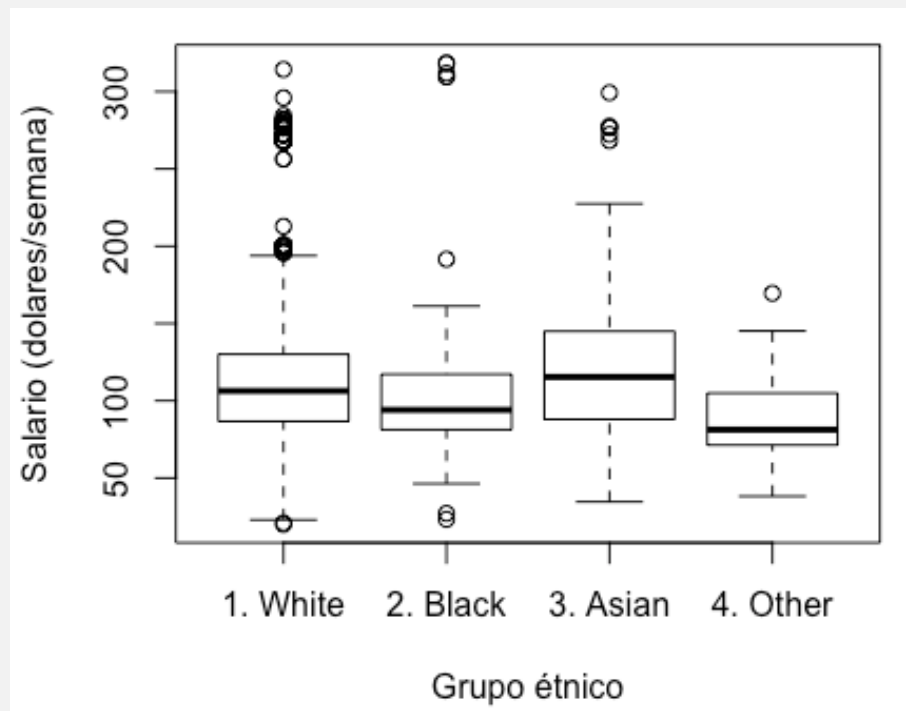
Elimina la columna X

Excluye todas las líneas con un salario mayor a 280 USD

Haz una gráfica de dispersión del salario en función de la edad y colorea los puntos por nivel de educación.

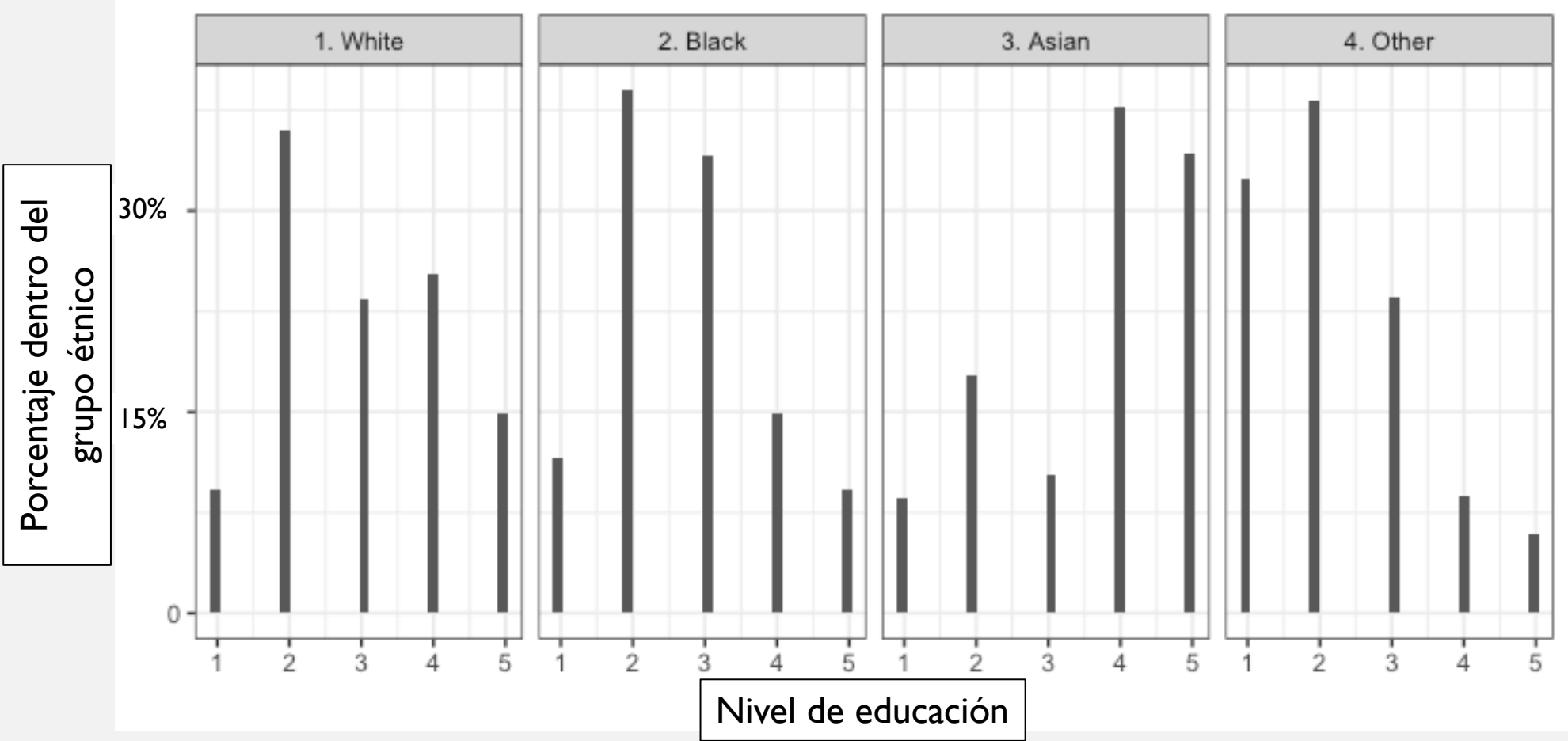
Cambia los títulos de los ejes.

## Ejercicio:



El salario incrementa de 16.33 dólares por cada nivel de educación

Ejercicio:



## Ejercicio en casa:

- Crear una variable numérica del nivel de educación usando:

```
Wage$edu.num <- as.numeric(Wage$education)
```

- Traducir el nombre de las columnas a Español
- Crear una función que corrija el salario de acuerdo al nivel de educación
  - 16.33 dólares **menos** por cada nivel de educación
- Crear nueva columna con el salario corregido
- Graficar el salario por grupo étnico
  - Título del gráfico
  - Títulos de los ejes
  - Colores por grupo étnico