

# Proyecto de Ciencia y Tecnología

Pablo Dinamarca

3/12/2020

# Contents

<b>1</b>	<b>Proyecto CyT</b>	<b>2</b>
1.1	Cargamos la base de datos . . . . .	2
1.2	Preparamos las variables . . . . .	2
1.3	Exploración Inicial . . . . .	3
1.4	Indice-h vs Citas . . . . .	6
1.5	Indice-h vs Edad . . . . .	13
1.6	Indice-h vs Edad y Citas . . . . .	15
1.7	Estudios . . . . .	16
1.8	Grado Universitario . . . . .	19
1.9	Afiliación . . . . .	20

# Chapter 1

## Proyecto CyT

### 1.1 Cargamos la base de datos

```
library(readxl)
library(tidyverse)
library(hrbrthemes)
setwd("C:/Users/user/Desktop/Universidad/Ciencias y Tecnologia/Proyecto de Ciencia y Tec
Data <- read_excel("Data.xlsx")
```

### 1.2 Preparamos las variables

```
Data$Genero <- as.factor(ifelse(Data$Femenino==1, "F", "M"))
Data$Titulo <- as.factor(ifelse(Data$Grado==1, "G",
                                ifelse(Data$Especialidad / Diplomado==1, "E",
                                        ifelse(Data$Maestría==1, "M",
                                              ifelse(Data$Doctorado=

# G = Grado
# E = Especialidad/Diplomado
# M = Master
# D = Doctorado
# PD = Post Doctorado

Data <- Data %>% select(-c("Post-Doctorado", "Doctorado", "Maestría",
                           "Especialidad / Diplomado", "Grado", "Femenino",
                           "Masculino"))
```

## 1.3 Exploración Inicial

```
nrow(Data)
```

```
## [1] 102
```

```
ncol(Data)
```

```
## [1] 23
```

Nuestro Dataset cuenta con 102 filas y 23 columnas.

```
table(Data$Genero)
```

```
##
```

```
##  F  M
```

```
## 19 83
```

```
data.frame(Promedio=mean(Data$Genero=="M")*100)
```

Promedio
81.37255

Notamos que existe una predominancia del genero masculino del 81,4%.

```
table(Data$Titulo)
```

```
##
```

```
##  D  M PD
```

```
## 81 13  8
```

```
mean(Data$Titulo=="D")*100
```

```
## [1] 79.41176
```

La predominancia del mayor nivel de estudios alcanzado es de la categoria de doctorados con un 79,4%.

```
table(Data$Nivel)
```

```
##
## Candidato a Investigador          I          II
##              57              33              8
##              III
##              4
```

Distribucion por nivel del Investigador

```
table(Data$Categoría)
```

```
##
## Activo Asociado
##      87      15
```

Distribucion por categoria del Investigador

```
mean(Data$`Año de obtención del ultimo grado académico`)
```

```
## [1] 2015.922
```

Promedio desde la obtencion del ultimo grado académico

```
mean(Data$`Años desde la Obtención del Último grado académico`)
```

```
## [1] 4.078431
```

Promedio de años desde la obtencion del ultimo grado académico

```
mean(Data$Edad)
```

```
## [1] 39.89216
```

Promedio de la edad de los investigadores

```
table(Data$`Estudió en el extranjero?`, Data$Titulo, dnn = c("Extranjero", "Titulo"))
```

```
##          Titulo
## Extranjero D  M PD
##          0 12  7  0
##          1 69  6  8
```

Distribucion de investigadores que estudiaron en el extranjero y su titulo obtenido

```
mean(Data$`Estudió en el extranjero?`)*100
```

```
## [1] 81.37255
```

Porcentaje que estudio en el extranjero

```
mean(Data$`Fue becado?`)*100
```

```
## [1] 82.35294
```

Porcentaje que fue becado

```
mean(Data$`Estudió en el extranjero?` & Data$`Fue becado?`)*100
```

```
## [1] 66.66667
```

Porcentaje que estudio en el extranjero y fue becado

```
table(Data$`Estudió en el extranjero?`, Data$`Fue becado?`, dnn = c("Extranjero", "Becado"))
```

```
##           Becado
## Extranjero 0  1
##           0  3 16
##           1 15 68
```

Distribucion de investigadores que estudiaron en el extranjero y fueron becados

```
mean(Data$`Trabaja en una Institución Pública?`)*100
```

```
## [1] 62.7451
```

Porcentaje que Trabaja en una Institución Pública

```
mean(Data$`Se hace investigación en el lugar donde trabaja?`)*100
```

```
## [1] 63.72549
```

Porcentaje donde Se hace investigación en el lugar donde trabaja

```
mean(Data$`Trabaja en una Institución Pública?` & Data$`Se hace investigación en el lugar donde trabaja`)*100
```

```
## [1] 43.13725
```

Porcentaje que Trabaja en una Institución Pública y Se hace investigación en el lugar donde trabaja

```
mean(Data$`Ejerce la docencia?`)*100
```

```
## [1] 85.29412
```

Porcentaje que Ejerce la docencia

```
mean(Data$`En una institución pública?`)*100
```

```
## [1] 78.43137
```

Porcentaje que Ejerce la docencia en una institución pública

```
mean(Data$`Participó en proyectos financiados por CONACYT`)*100
```

```
## [1] 57.84314
```

Porcentaje que Participó en proyectos financiados por CONACYT

```
mean(Data$`Participó en proyectos financiados por organismos internacionales`)*100
```

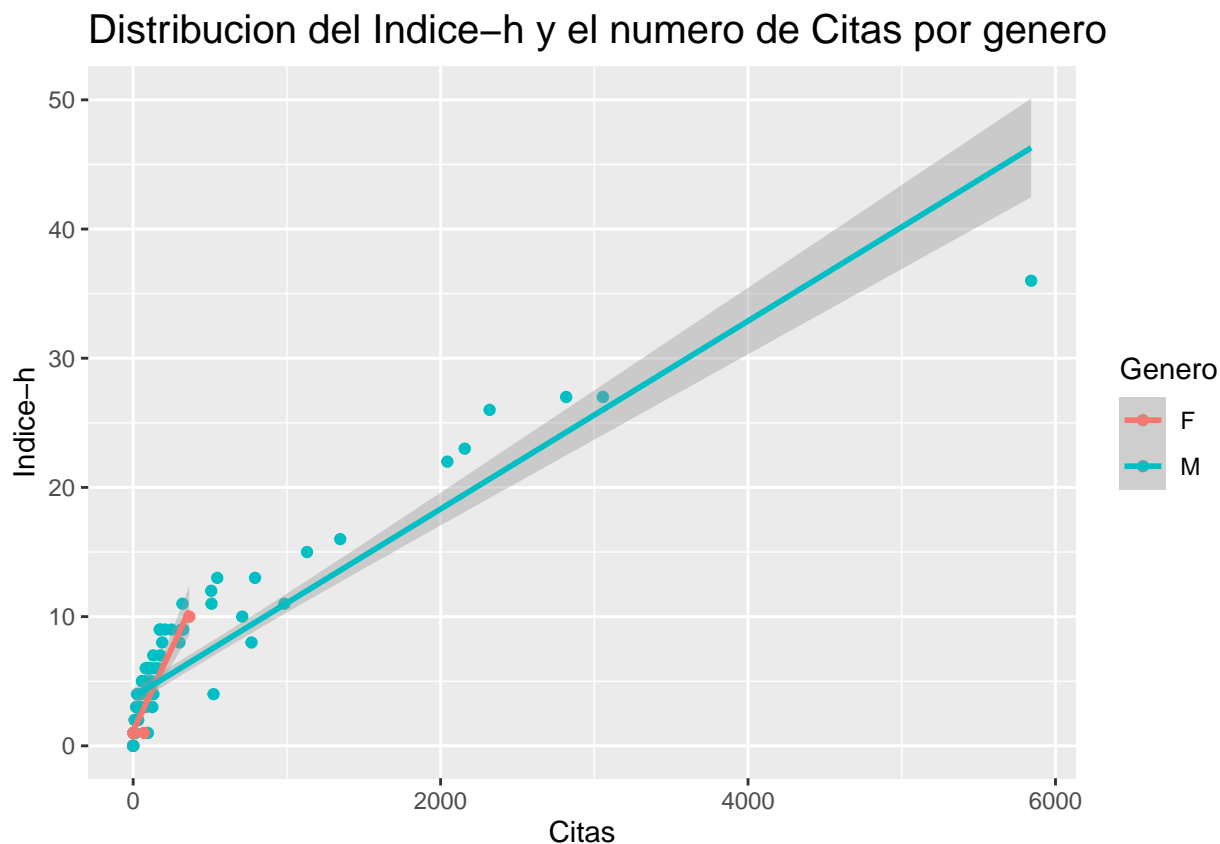
```
## [1] 56.86275
```

Porcentaje que Participó en proyectos financiados por organismos internacionales

## 1.4 Índice-h vs Citas

```
Data %>% ggplot(aes(Citas, `Índice-h`, color=Genero)) + geom_point() + geom_smooth(method="lm") +  
  ggtitle("Distribucion del Índice-h y el numero de Citas por genero") +  
  theme(plot.title = element_text(size=15))
```

```
## `geom_smooth()` using formula 'y ~ x'
```



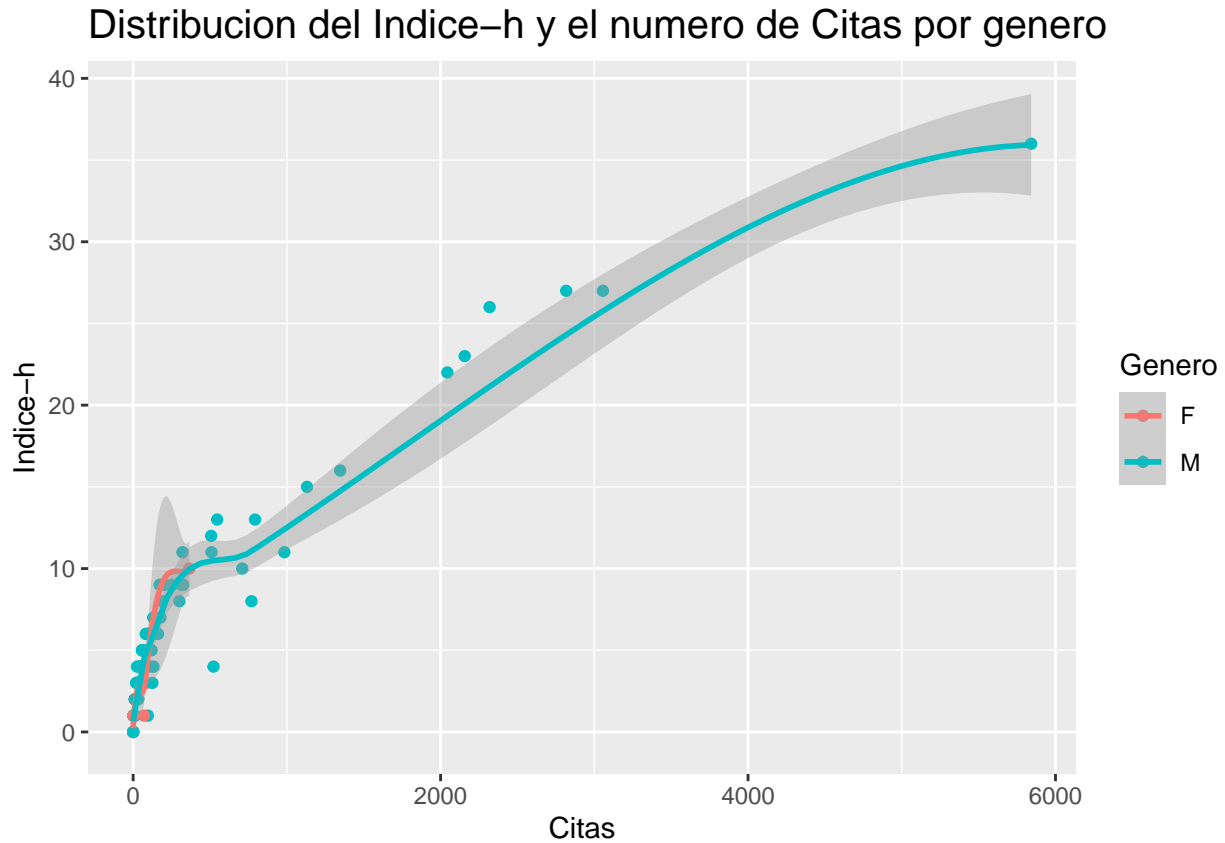
En este grafico podemos observar la distribucion del Indice-h vs el numero de citas con sus respectivas tendencias clasificadas por sexo.

Notamos una correlacion positiva y una pendiente mas elastica para los hombres. Esto quiere decir que ante mayor numero de citas, mayor es tambien el indice-h.

```
Data %>% ggplot(aes(Citas, `Indice-h`, color=Genero)) + geom_point() + geom_smooth() +  
  ggtitle("Distribucion del Indice-h y el numero de Citas por genero") +  
  theme(plot.title = element_text(size=15))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

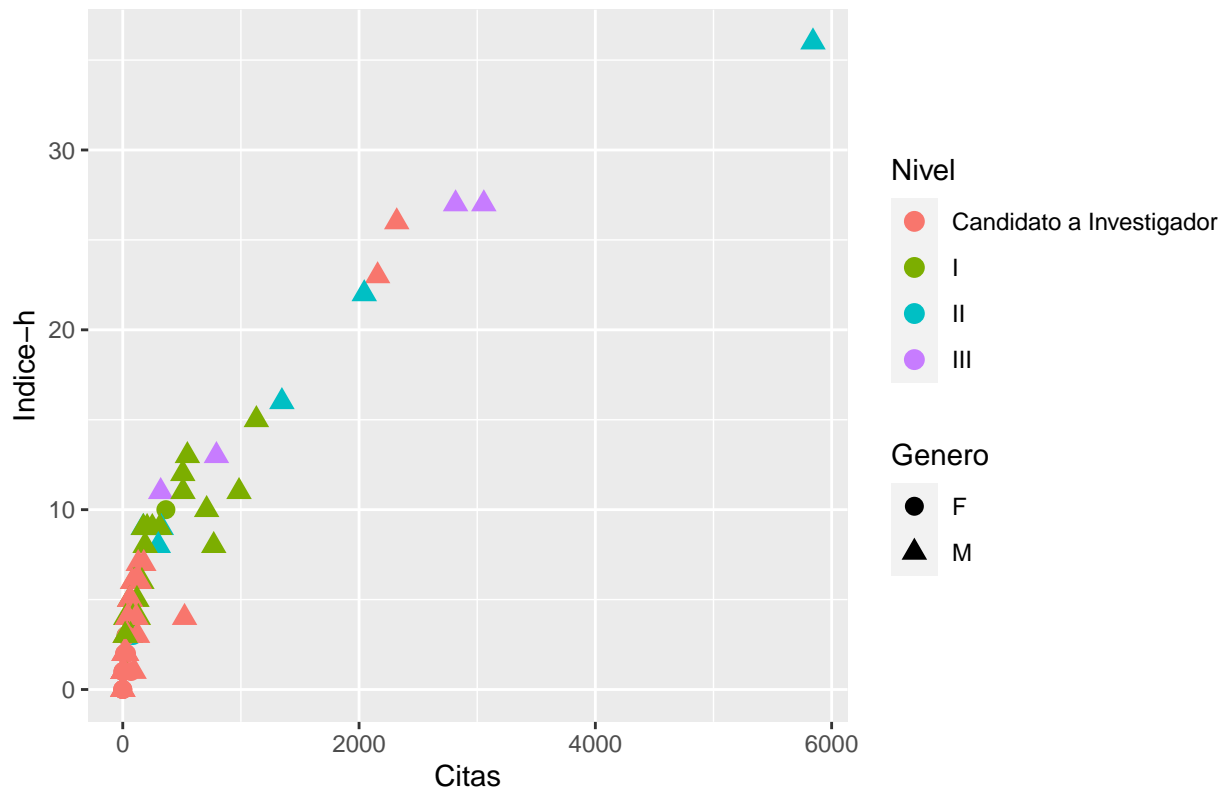




Este grafico nos señala una tendencia mas precisa de los datos con el respectivo intervalo de confianza del 95%.

```
Data %>% ggplot(aes(Citas, `Indice-h`, color=Genero)) + geom_point(aes(shape=Genero), size=100) +
  ggtitle("Distribucion del Indice-h y el numero de Citas por nivel y genero") +
  theme(plot.title = element_text(size=15))
```

## Distribucion del Indice-h y el numero de Citas por nivel y genero

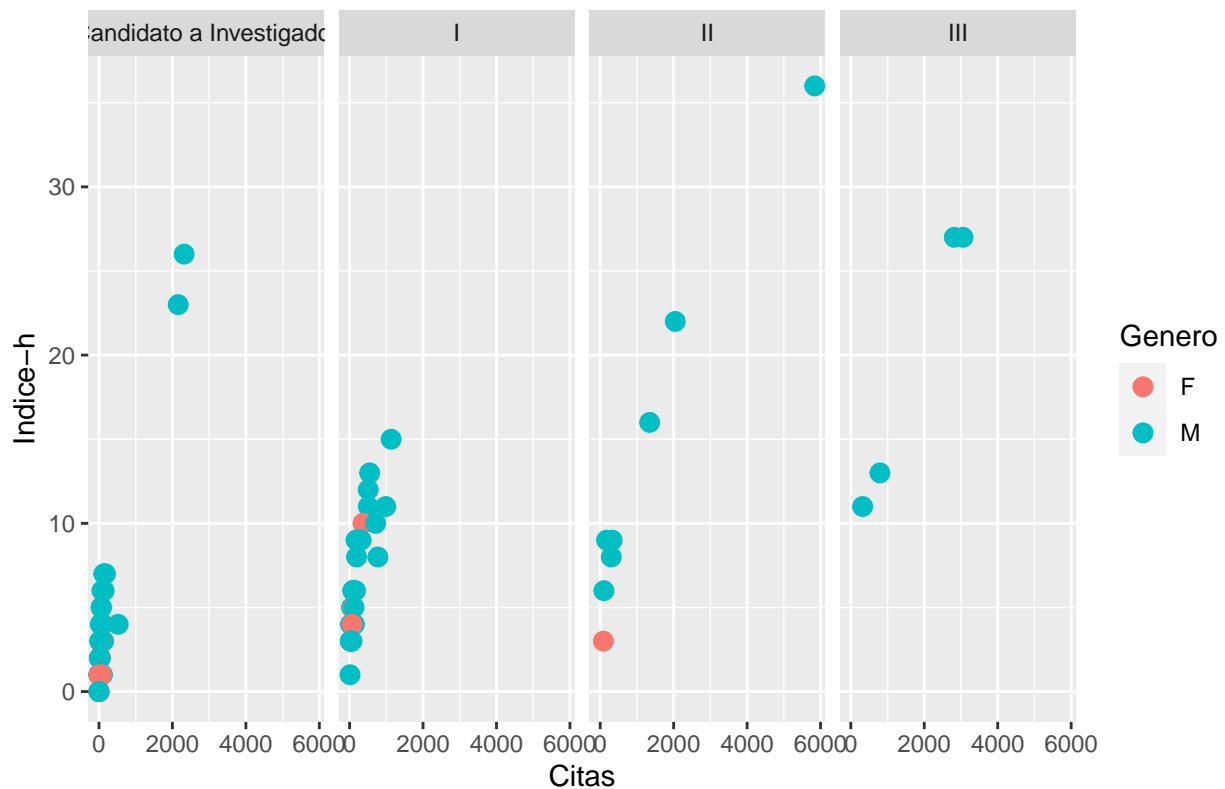


Con este grafico notamos la clasificacion por colores del nivel de los investigadores.

Podemos notar que un investigador de nivel II tiene mayo numero de citas tanto como de indice-h. Ademas se diferencia en el grafico los sexos por formas.

```
Data %>% ggplot(aes(Citas, `Indice-h`, color=Genero)) + geom_point(size=3) + facet_grid(  
  ggtitle("Distribucion del Indice-h y el numero de Citas por nivel y genero") +  
  theme(plot.title = element_text(size=15))
```

## Distribucion del Indice-h y el numero de Citas por nivel y genero

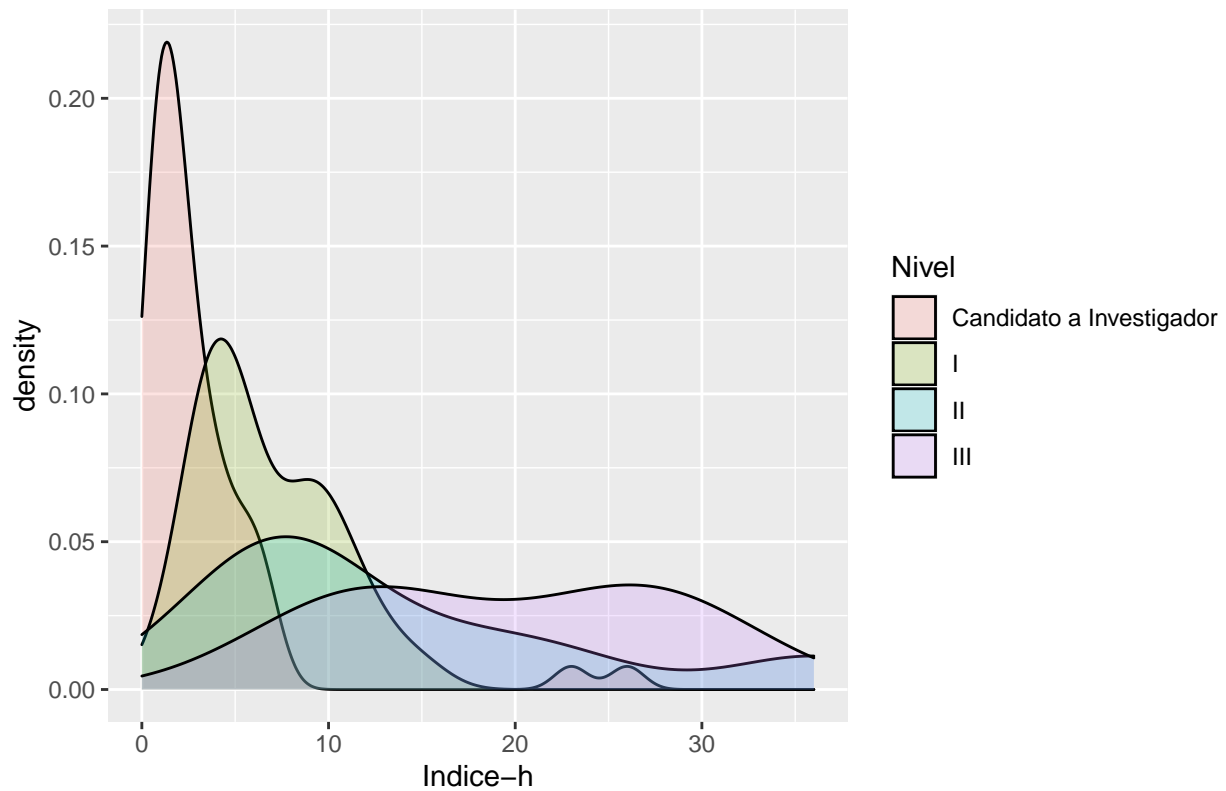


En el grafico, se hace mas evidente la distribucion de las variables por nivel de investigador.

Notamos ue el nivel I cuenta con menos cantidad de citas e indice-h mientras que el nivel II es lo contrario.

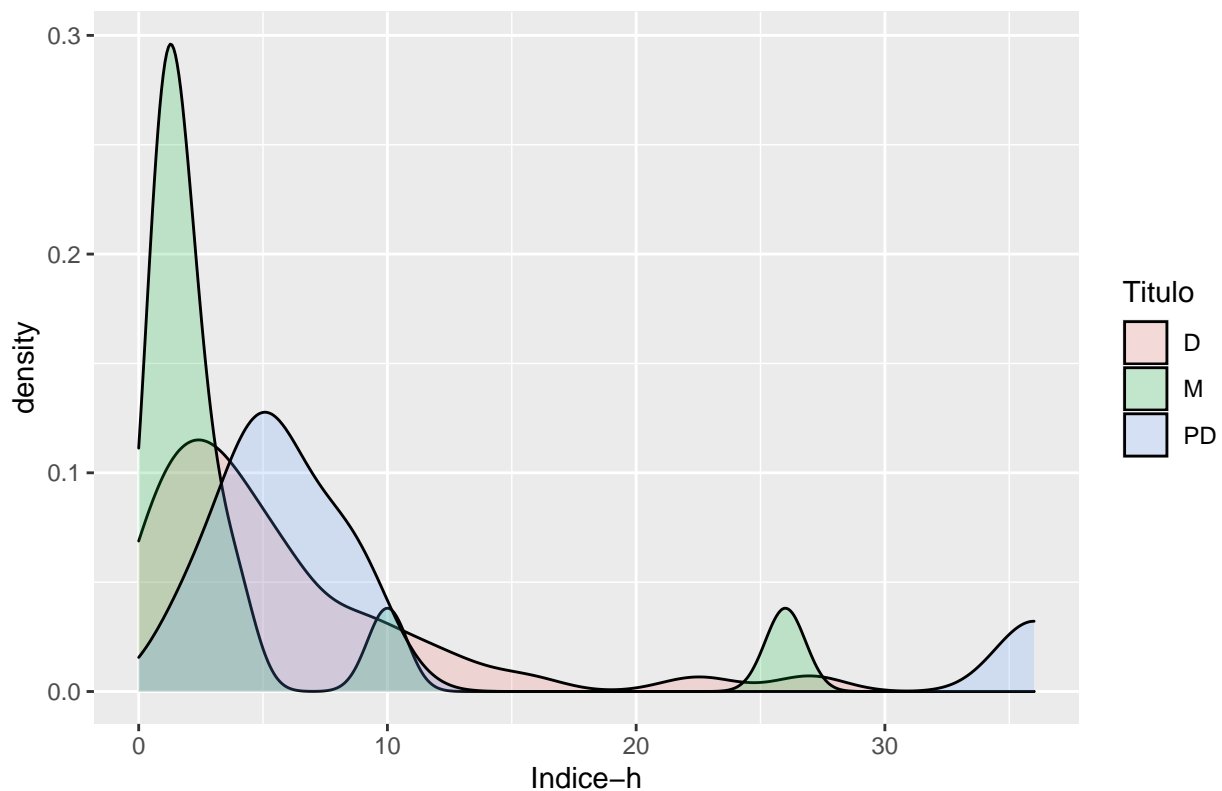
```
Data %>% ggplot(aes(`Indice-h`, fill=Nivel)) + geom_density(alpha=0.2) +  
  ggtitle("Distribucion del Indice-h y el numero de Citas por nivel y genero") +  
  theme(plot.title = element_text(size=15))
```

## Distribucion del Indice-h y el numero de Citas por nivel y genero



```
Data %>% ggplot(aes(`Indice-h`, fill=Titulo)) + geom_density(alpha=0.2) +  
  ggtitle("Distribucion del Indice-h y el numero de Citas por nivel y genero") +  
  theme(plot.title = element_text(size=15))
```

## Distribucion del Indice-h y el numero de Citas por nivel y genero



```
Data %>% ggplot(aes(`Indice-h`, fill=Nivel)) + geom_density(alpha=0.2) +
  facet_grid(Titulo~Genero)+
  ggtitle("Distribucion del Indice-h y el numero de Citas por nivel y genero") +
  theme(plot.title = element_text(size=15))
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

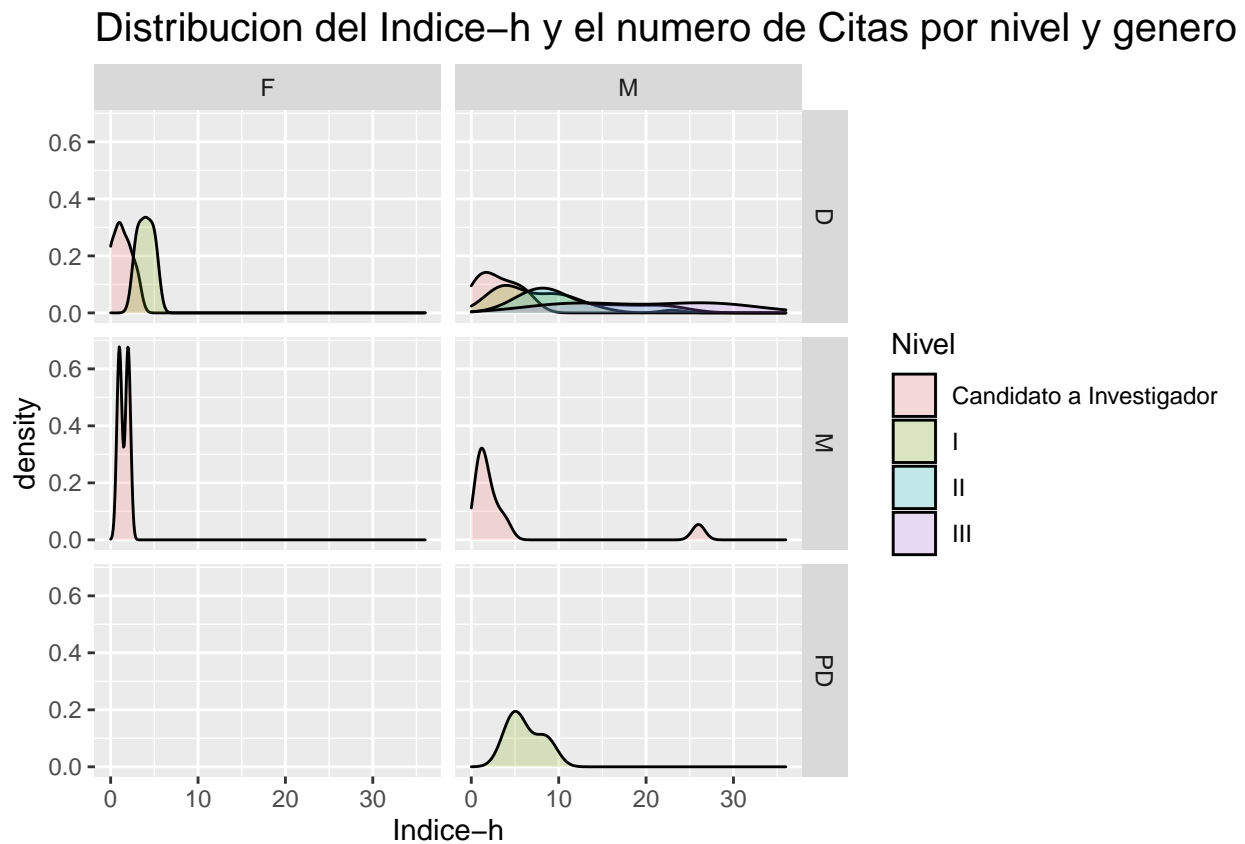
```
## Warning in max(ids, na.rm = TRUE): ningun argumento finito para max; retornando
## -Inf
```

```
## Warning in max(ids, na.rm = TRUE): ningun argumento finito para max; retornando
## -Inf
```

```
## Warning in max(ids, na.rm = TRUE): ningun argumento finito para max; retornando
```

```
## -Inf
```

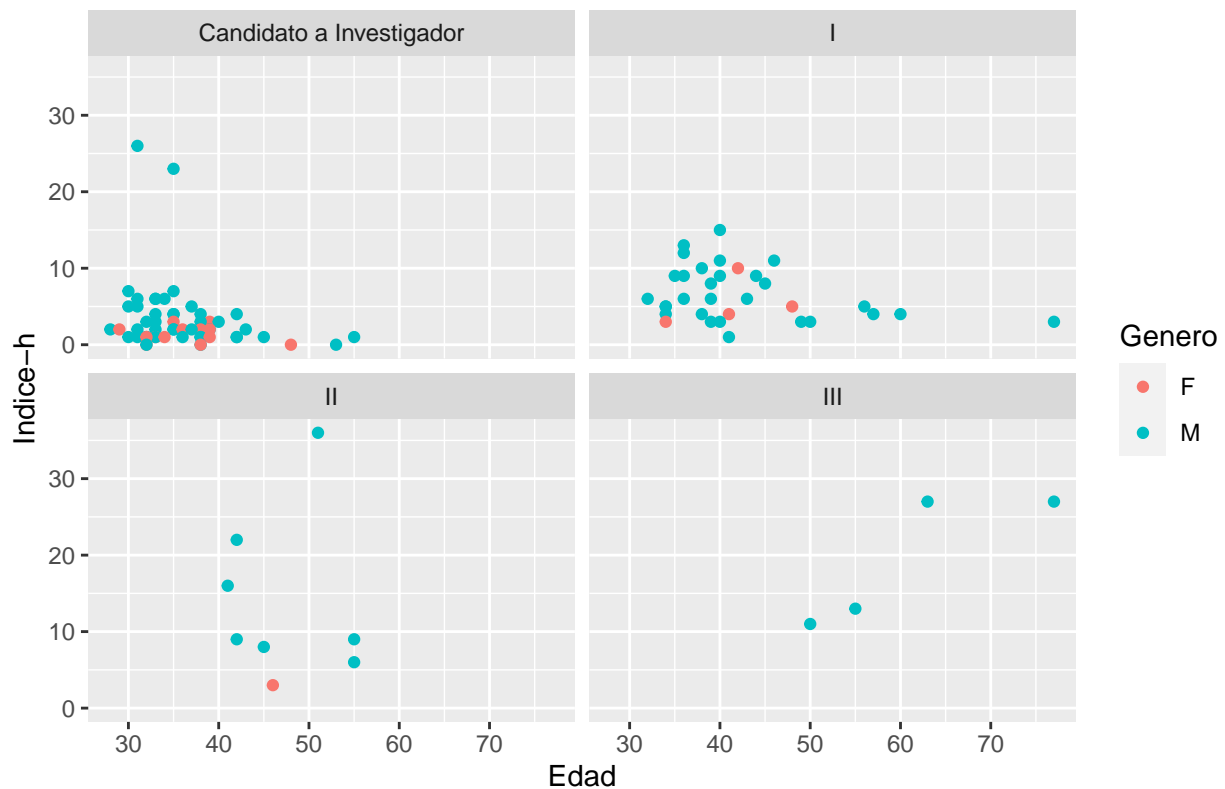
```
## Warning in max(ids, na.rm = TRUE): ningun argumento finito para max; retornando  
## -Inf
```



## 1.5 Indice-h vs Edad

```
Data %>% ggplot(aes(Edad, `Indice-h`, color=Genero)) + geom_point() + facet_wrap(~Nivel)  
  ggtitle("Distribución del Indice-h vs la Edad del Investigador") +  
  theme(plot.title = element_text(size=15))
```

## Distribución del Índice-h vs la Edad del Investigador

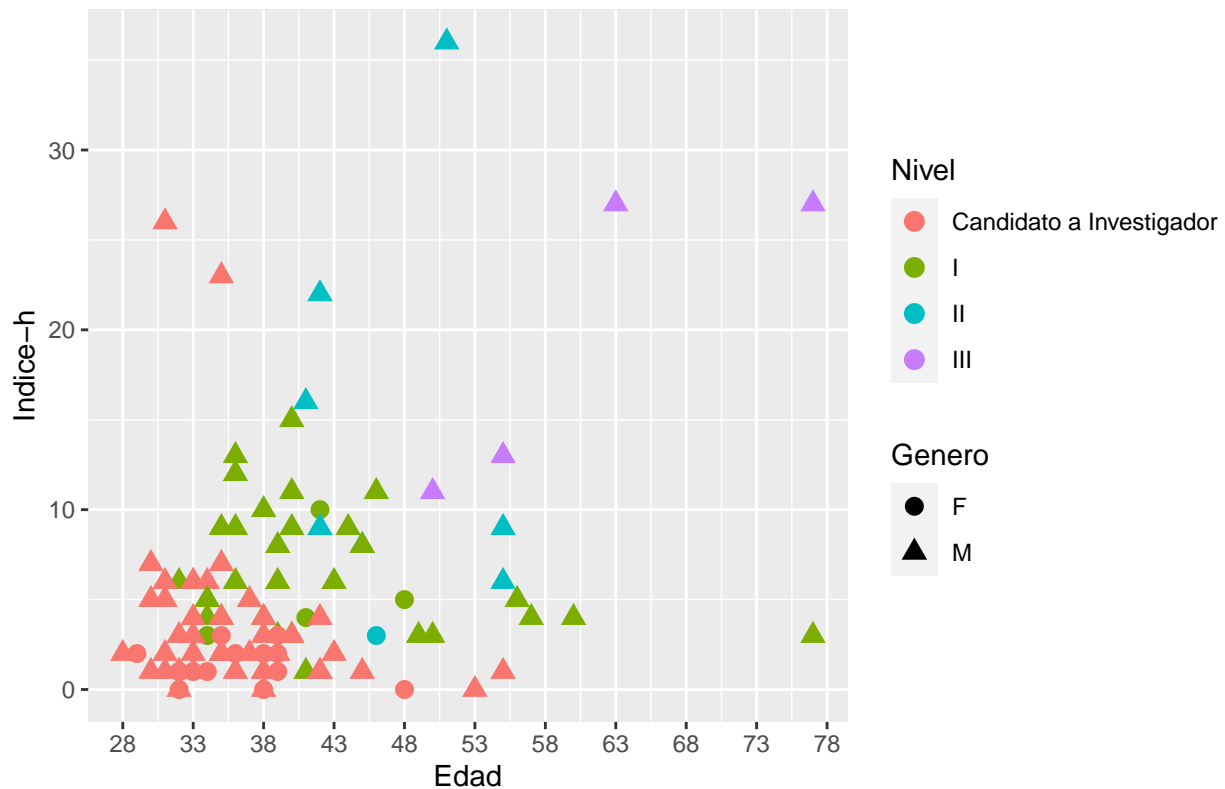


En el presente grafico tenemos la distribucion por edad vs el indice-h dividido por el nivel del investigador y clasificado por sexo.

Notamos que para los candidatos a investigador, los valores se centran mas entre los 30 y 45 años de edad similar a los de nivel I. mientras que para los de nivel II, el rango esta entre los 40 y 55 años con un indice-h promedio superior.

```
Data %>% ggplot(aes(Edad, `Índice-h`, color=Nivel)) + geom_point(aes(shape=Genero), size=
  scale_x_continuous(breaks=seq(28, 100, 5)) +
  ggtitle("Distribución del Índice-h vs la Edad del Investigador") +
  theme(plot.title = element_text(size=15))
```

## Distribución del Índice-h vs la Edad del Investigador



Este grafico nos presenta el indice-h vs la edad clasificado por el nivel y genero del investigador.

Notamos una predominancia de los investigadores entre los 28 y 48 años y un indice-h de maximo 15.

## 1.6 Índice-h vs Edad y Citas

```
Data %>% ggplot(aes(Edad, `Índice-h`, color=Nivel)) +
  geom_point(aes(size=Citas), alpha=1) +
  geom_text(aes(label=Citas), nudge_x = 1.5) +
  scale_x_continuous(breaks=seq(28, 100, 5)) +
  ggtitle("Distribucion del Índice-h vs Edad y numero de Citas") +
  theme(plot.title = element_text(size=15))
```



## Distribucion del Indice-h vs Edad y numero de Citas

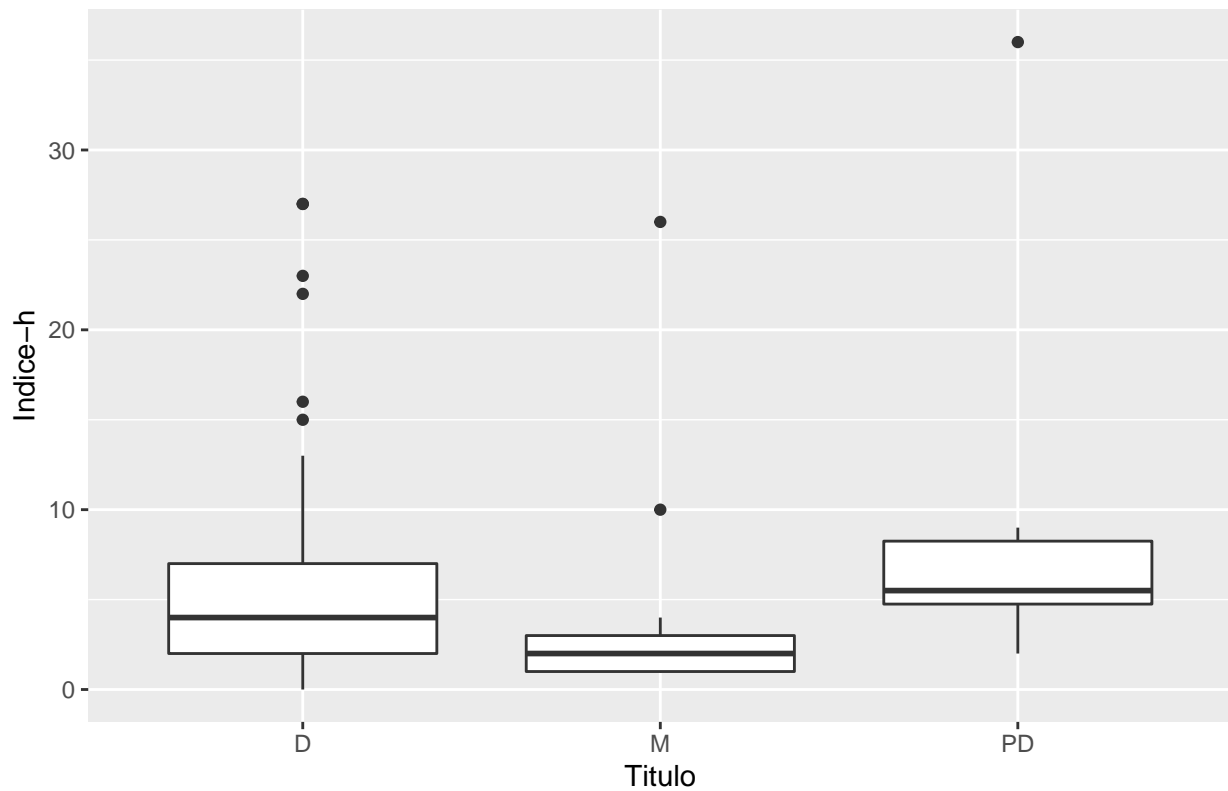


El presente grafico nos muestra la clasificacion de los investigadores por nivel y numero de citas.

## 1.7 Estudios

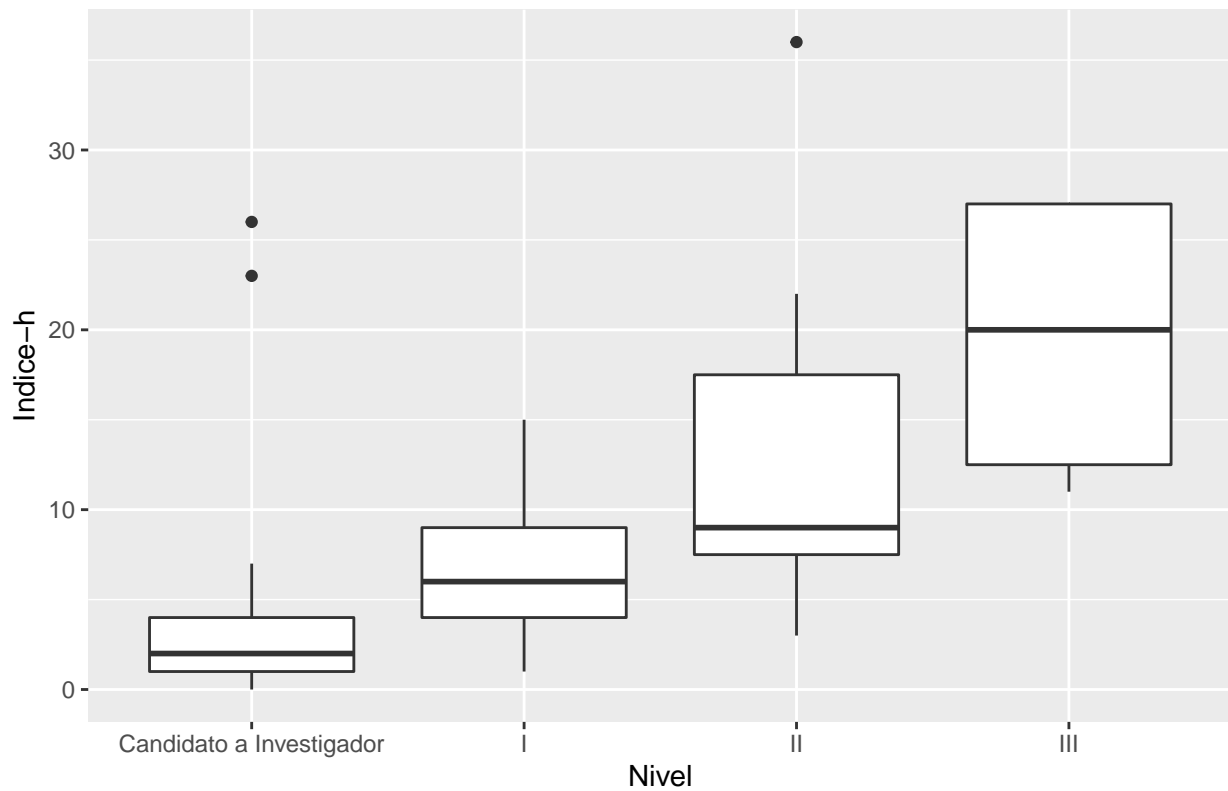
```
Data %>% ggplot(aes(Titulo, `Indice-h`)) +
  geom_boxplot() +
  ggtitle("Distribución del Indice-h vs la Edad del Investigador") +
  theme(plot.title = element_text(size=15))
```

## Distribución del Índice-h vs la Edad del Investigador



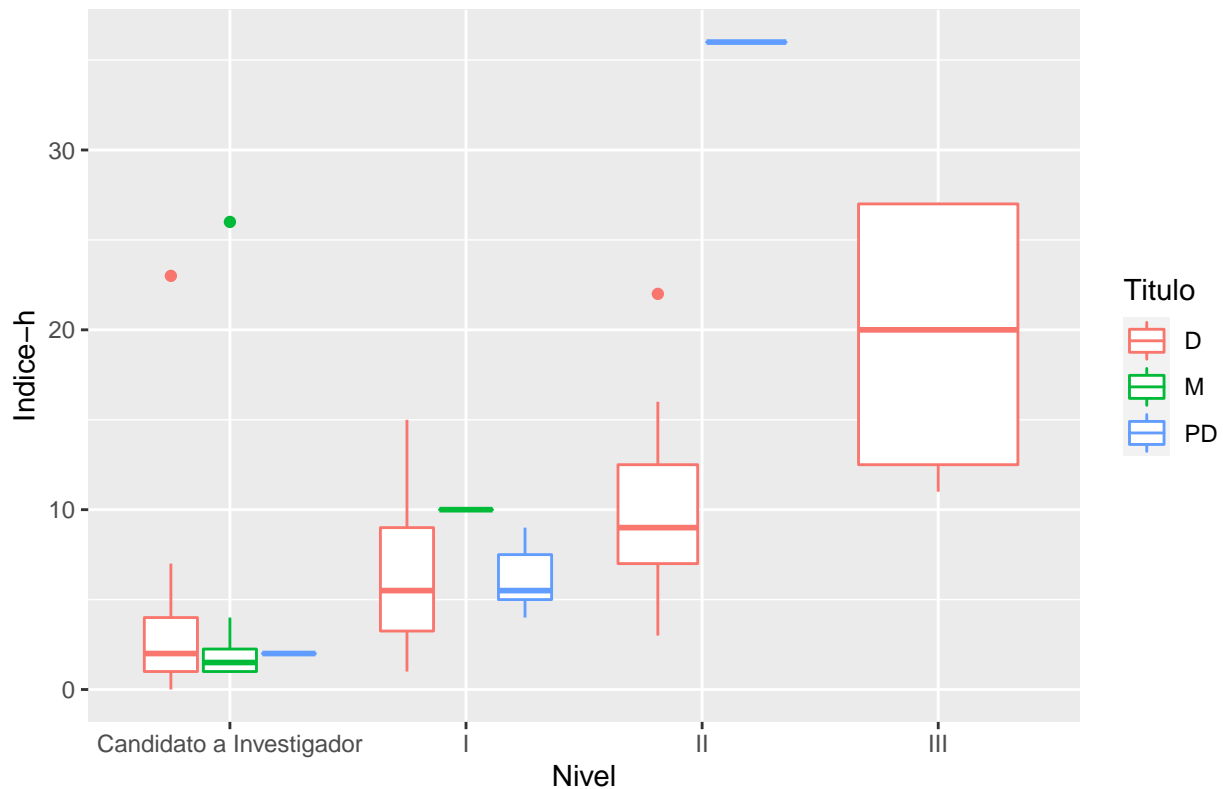
```
Data %>% ggplot(aes(Nivel, `Índice-h`)) +  
  geom_boxplot() +  
  ggtitle("Distribución del Índice-h vs la Edad del Investigador") +  
  theme(plot.title = element_text(size=15))
```

## Distribución del Índice-h vs la Edad del Investigador



```
Data %>% ggplot(aes(Nivel, `Índice-h`, color=Titulo)) +  
  geom_boxplot() +  
  ggtitle("Distribución del Índice-h vs la Edad del Investigador") +  
  theme(plot.title = element_text(size=15))
```

## Distribución del Índice-h vs la Edad del Investigador

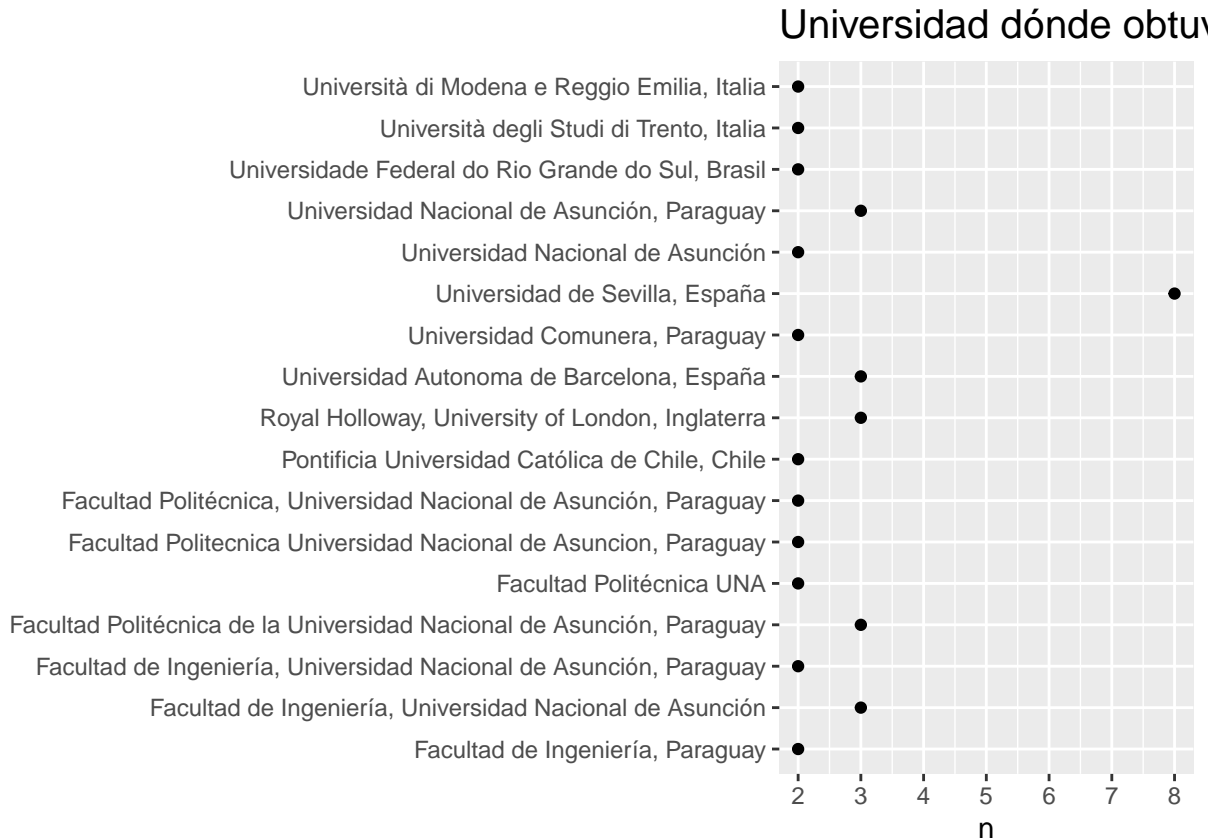


## 1.8 Grado Universitario

```
Uni <- Data %>% select(`Universidad dónde obtuvo su último grado universitario`) %>%
  group_by(`Universidad dónde obtuvo su último grado universitario`) %>% summarise(n=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Uni %>% filter(n>1) %>% ggplot(aes(n, `Universidad dónde obtuvo su último grado universitario`)) +
  geom_point() + scale_x_continuous(breaks = seq(0, 8)) +
  ylab("") +
  ggtitle("Universidad dónde obtuvo su último grado universitario") +
  theme(plot.title = element_text(size=15))
```



En el grafico podemos notar que algunos investigadores realizaron sus estudios en universidades similares.

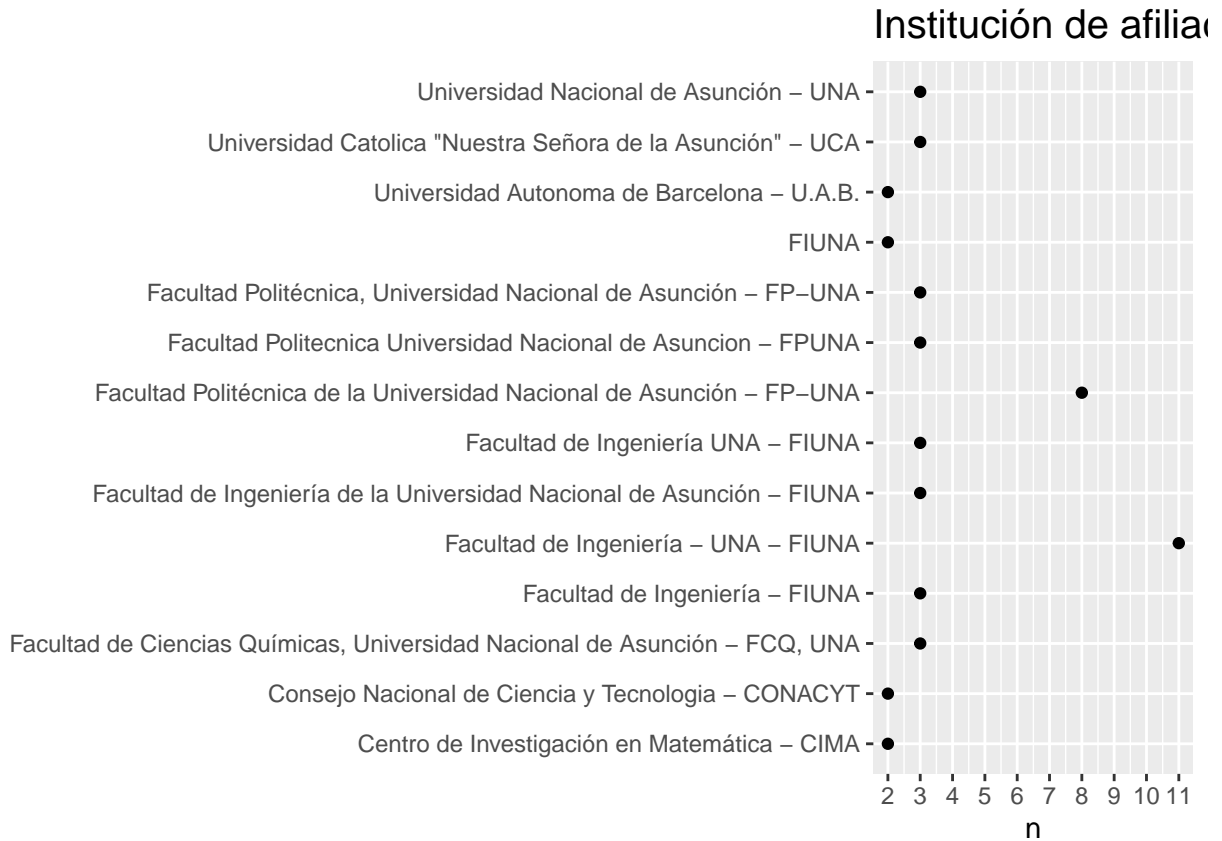
Se destaca la Universidad de Sevilla, España y la UNA

## 1.9 Afiliación

```
Afi <- Data %>% select(`Institución de afiliacion`) %>%
  group_by(`Institución de afiliacion`) %>% summarise(n=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Afi %>% filter(n>1) %>% ggplot(aes(n, `Institución de afiliacion`)) +
  geom_point() + scale_x_continuous(breaks = seq(0, 11)) +
  ylab("") +
  ggtitle("Institución de afiliacion") +
  theme(plot.title = element_text(size=15))
```



En el gráfico podemos notar que algunos investigadores están afiliados a Instituciones similares.

Se encuentra una principal predominancia de diferentes sectores de la UNA