

Competitividad regional en la Unión Europea

1.

La base de datos proporcionada de la Comisión Europea tiene las siguientes características:

- **Dimensión:** 234 filas y 21 columnas.
- **Columnas:**
 - **REGION:** Nombre de la región (tipo objeto, texto).
 - **NUTS2ID:** Código NUTS2 de la región (objeto, texto).
 - **CNTR:** Código del país (objeto, texto).
 - **DEVLEV:** Nivel de desarrollo de la región (objeto, texto).
 - **RANK:** Rango de la región (entero).
 - **RCI, BSI, INSTIT, MACRO, INFRASTR, HEALTH, BASICEDUC, ESI, HIGHEREDUC, LABMARKET, MARKETSIZE, ISI, TECHREAD, BUSINESSSOPH, INNOV, GDPPC:** Estas variables son valores numéricos que miden distintos indicadores como Competitividad Regional (RCI), Infraestructura (INFRASTR), Mercado Laboral (LABMARKET), PIB per cápita (GDPPC), entre otros.

2.

El análisis descriptivo de las variables cuantitativas del conjunto de datos muestra lo siguiente:

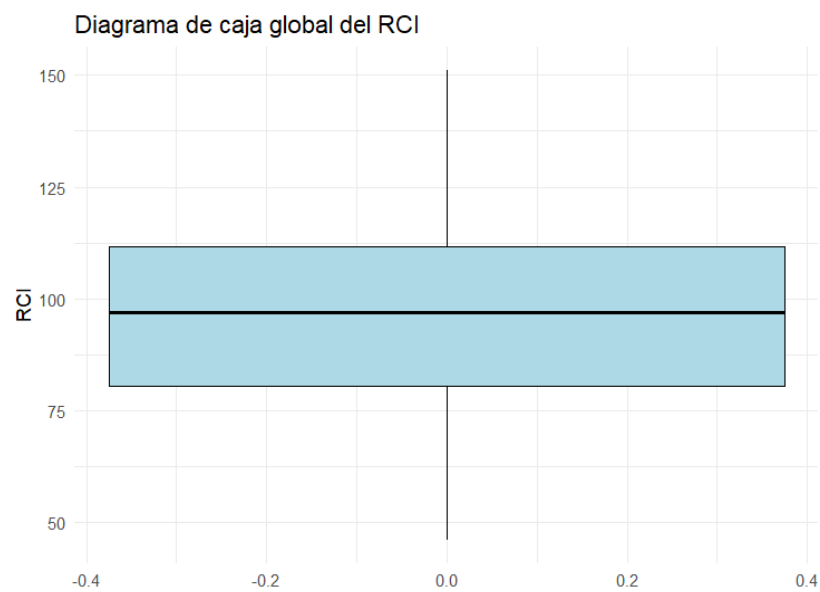
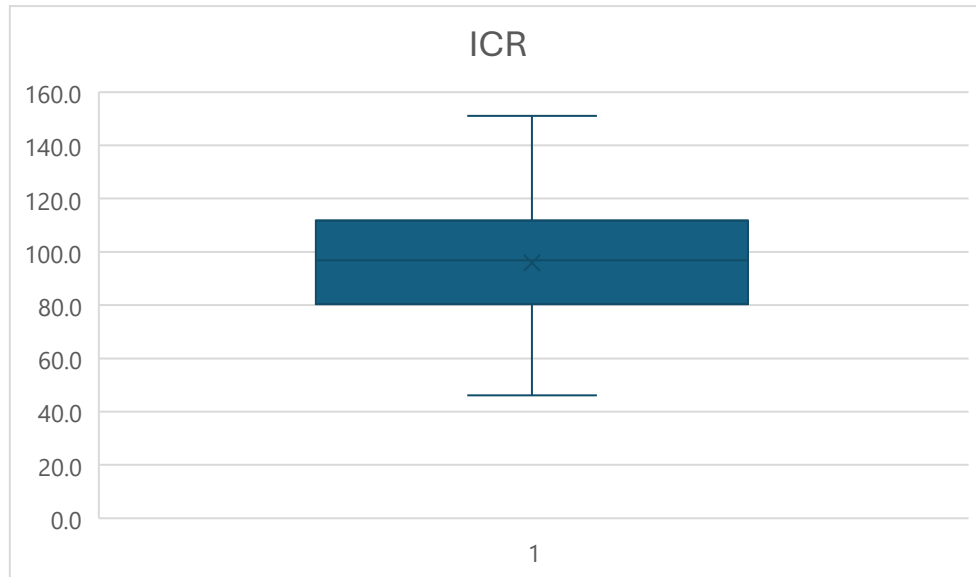
- **RANK:** Rango de la región.
 - Media: 117.35
 - Mínimo: 1, Máximo: 234
 - Percentil 25: 59.25, Percentil 50: 117.5, Percentil 75: 175.75
- **RCI** (Regional Competitiveness Index):
 - Media: 95.9

- Mínimo: 46.1, Máximo: 151.1
- **BSI** (Basic Subindex):
 - Media: 97.37
 - Mínimo: 37.3, Máximo: 138.1
- **INSTIT** (Institutions):
 - Media: 102.53
 - Mínimo: 34.1, Máximo: 197.9
- **MACRO** (Macroeconomic stability):
 - Media: 101.45
 - Mínimo: 33.6, Máximo: 159.7
- **INFRASTR** (Infrastructure):
 - Media: 85.82
 - Mínimo: 19.5, Máximo: 185.8
- **HEALTH** (Health):
 - Media: 97.05
 - Mínimo: 41.0, Máximo: 127.2
- **BASICEDUC** (Basic Education):
 - Media: 98.64
 - Mínimo: 3.7, Máximo: 152.3
- **ESI** (Efficiency Subindex):
 - Media: 94.94
 - Mínimo: 40.1, Máximo: 157.4
- **HIGHEREDUC** (Higher Education):
 - Media: 97.83
 - Mínimo: 27.7, Máximo: 155.4
- **LABMARKET** (Labor Market):
 - Media: 99.05
 - Mínimo: 12.0, Máximo: 126.8
- **MARKETSIZE** (Market Size):
 - Media: 80.99
 - Mínimo: 3.9, Máximo: 274.6

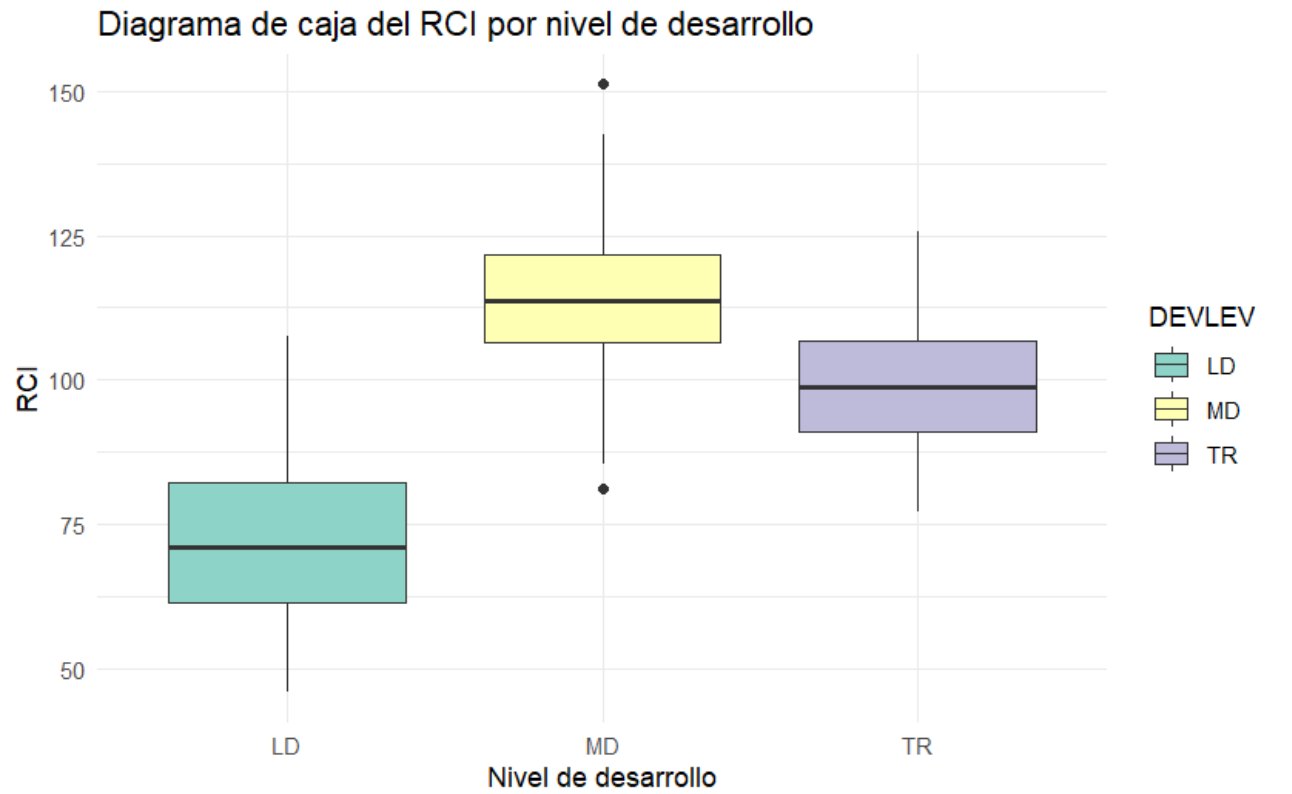
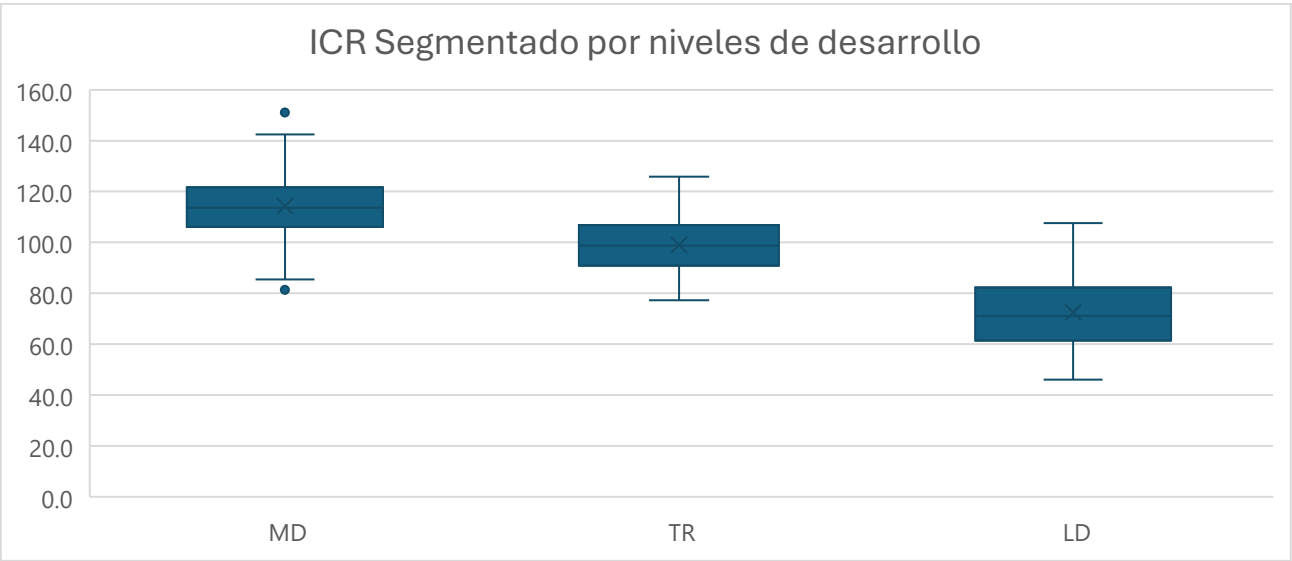
- **ISI** (Innovation Subindex):
 - Media: 95.68
 - Mínimo: 23.7, Máximo: 166.9
- **TECHREAD** (Technological Readiness):
 - Media: 99.09
 - Mínimo: 13.1, Máximo: 175.7
- **BUSINESSSOPH** (Business Sophistication):
 - Media: 94.76
 - Mínimo: 12.5, Máximo: 184.4
- **INNOV** (Innovation):
 - Media: 92.16
 - Mínimo: 12.9, Máximo: 193.8
- **GDPPC** (GDP per capita):
 - Media: 93.41
 - Mínimo: 28.7, Máximo: 259.63

3.

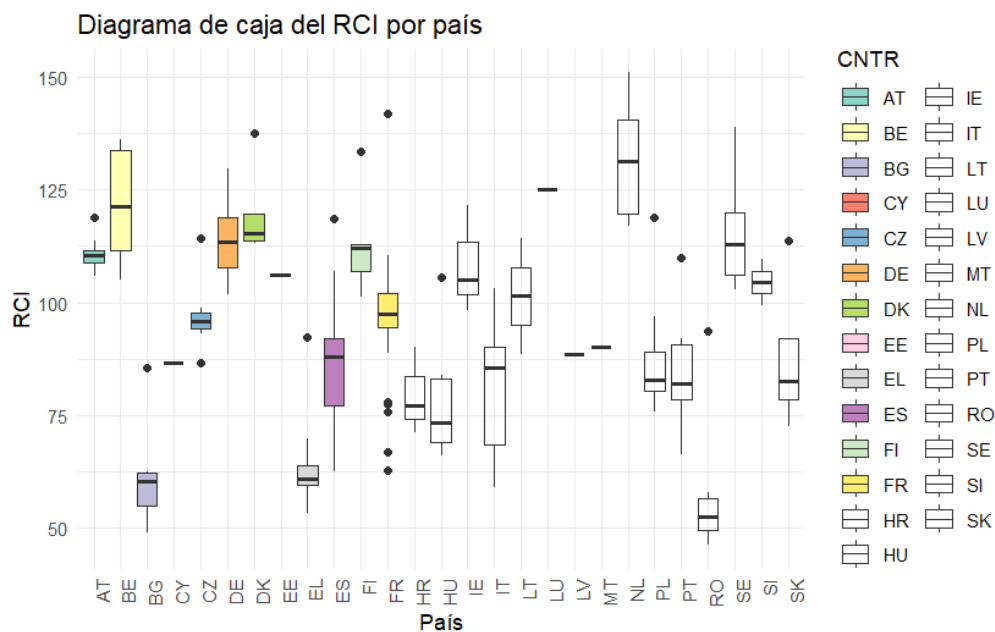
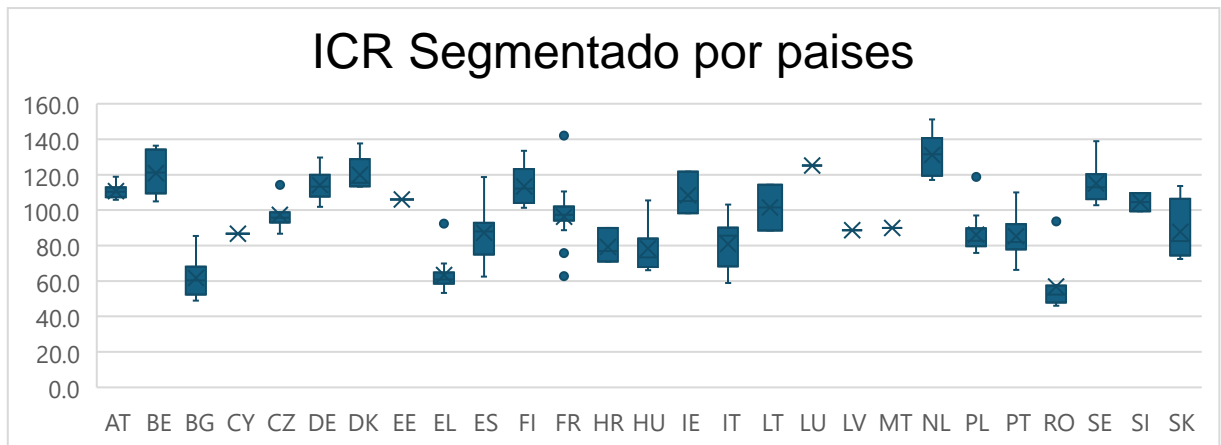
Se muestra los diagramas de caja del ICR generados tanto por RStudio como por Excel:



A continuación, se muestran los diagramas de caja del ICR segmentados por región generados tanto por RStudio como por Excel:



A continuación, se muestran los diagramas de caja del ICR segmentados por países generados tanto por RStudio como por Excel:



Los respectivos códigos de RStudio desarrollados para llevar a cabo los diagramas son:

```
# Boxplot global del RCI
ggplot(datos, aes(y = RCI)) +
  geom_boxplot(fill = "lightblue", color = "black") +
  labs(title = "Diagrama de caja global del RCI", y = "RCI") +
  theme_minimal()

# Boxplot del RCI condicionado por el nivel de desarrollo
ggplot(datos, aes(x = DEVLEV, y = RCI, fill = DEVLEV)) +
  geom_boxplot() +
  labs(title = "Diagrama de caja del RCI por nivel de desarrollo", x =
"Nivel de desarrollo", y = "RCI") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3")

# Boxplot del RCI condicionado por país
ggplot(datos, aes(x = CNTR, y = RCI, fill = CNTR)) +
  geom_boxplot() +
```

```
labs(title = "Diagrama de caja del RCI por país", x = "País", y = "RCI") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 90, hjust = 1)) + # Rotar
etiquetas de país
scale_fill_brewer(palette = "Set3")
```

4.

10 mayores valores de ICR por país:

Código RStudio:

```
# Obtener Los 10 mayores valores de ICR por país
top_10_RCI_pais <- datos %>%
  group_by(CNTR) %>%
  arrange(desc(RCI)) %>%
  slice_head(n = 10)
```

Excel con los resultados:



top_10_RCI_pais.csv

10 mayores valores de ICR por región:

```
# Obtener Los 10 mayores valores de ICR por región
top_10_RCI_region <- datos %>%
  group_by(NUTS2ID) %>%
  arrange(desc(RCI)) %>%
  slice_head(n = 10)
```

Excel con los resultados:



top_10_RCI_region.csv

v

10 menores valores de ICR por país:

Código RStudio:

```
# Obtener Los 10 menores valores de ICR por país  
bottom_10_RCI_pais <- datos %>%  
  group_by(CNTR) %>%  
  arrange(RCI) %>%  
  slice_head(n = 10)
```

Excel con los resultados:



bottom_10_RCI_pais.csv
sv

10 menores valores de ICR por región:

```
# Obtener Los 10 menores valores de ICR por región  
bottom_10_RCI_region <- datos %>%  
  group_by(NUTS2ID) %>%  
  arrange(RCI) %>%  
  slice_head(n = 10)
```

Excel con los resultados:



bottom_10_RCI_region
n.csv

5.

El archivo es de tipo GeoJSON, este está estructurado en función de una serie de características:

Tiene una geometría MultiPolygon, es decir, las regiones están formadas por múltiples áreas discontinuas

Presenta unos atributos los cuales muestran el nombre de la región, el código NUTS (que identifica únicamente cada región) y una serie de coordenadas asociada a cada código

El código utilizado:

```
# Cargar el archivo GeoJSON
mapa_nuts2 <- st_read("C:/Users/November/Downloads/EU_NUTS2.geojson")

# Ver Las primeras filas del dataset
head(mapa_nuts2)

# Ver La proyección (sistema de coordenadas)
st_crs(mapa_nuts2)

# Ver el tipo de geometría (polígonos)
st_geometry_type(mapa_nuts2)

# Resumen general del dataset
summary(mapa_nuts2)
```

6.

Para fusionar los ficheros EU_NUTS2.geojson y RCI_2_0_scores.xlsx utilizaremos el siguiente código de R:

Cargar el archivo GeoJSON

```
europe_map <- st_read("C:/Users/November/Downloads/EU_NUTS2.geojson")
```

```
# Cargar el archivo RCI_2_0
ruta_archivo <- "C:/Users/November/Downloads/RCI_2_0_scores.xlsx"

# Revisar Las columnas del mapa
colnames(europe_map)

## [1] "NUTS2ID" "AGGCOUNT" "geometry"
```

```
# Revisar Las columnas de Los datos de La Comisión Europea
colnames(datos)

Merged_data <- europe_map %>%
  left_join(datos, by = c("NUTS2ID" = "NUTS2ID"))

# Ver Las primeras filas de Los datos fusionados
head(merged_data)
```

Una vez hemos juntado los dos grupos de datos procedemos a estudiar sus características:

Este nuevo conjunto de datos tiene una clase sf y data.frame, es decir, contiene datos geoespaciales y datos numéricos.

La geometría de cada observación es de tipo Multipolygon, es decir, cada región está formada por más de un lugar. Podemos encontrar una gran cantidad de variables en este conjunto de datos:

- **NUTS2ID:** Código de identificación de las regiones NUTS 2 (formato de texto). Ejemplos: "AL01", "AL02", "AT11", etc.
- **AGGCOUNT:** Número de regiones agregadas. La mayoría de los valores son 1, lo que indica que no hay agregación significativa.
- **REGION:** Nombre de la región, aunque muchos están vacíos (NA).
- **CNTR:** Código de país (formato de texto, por ejemplo, "AT" para Austria), pero hay varios NA.
- **DEVLEV:** Nivel de desarrollo regional, con valores como "TR" (transición) y algunos NA.
- **RANK:** Rango del Índice de Competitividad Regional (ICR) de las regiones. Tiene algunos valores NA.

Variables socioeconómicas: Incluyen diversos índices y puntajes sobre competitividad, infraestructura, educación, etc. Ejemplos:

- **RCI:** Índice de Competitividad Regional (valor medio: 96.37).
- **GDPPC:** PIB per cápita (valor medio: 94.21).
- **INFRASTR, HEALTH, BASICEDUC, LABMARKET, INNOV,** entre otros.

El Código utilizado para obtener la información es el siguiente:

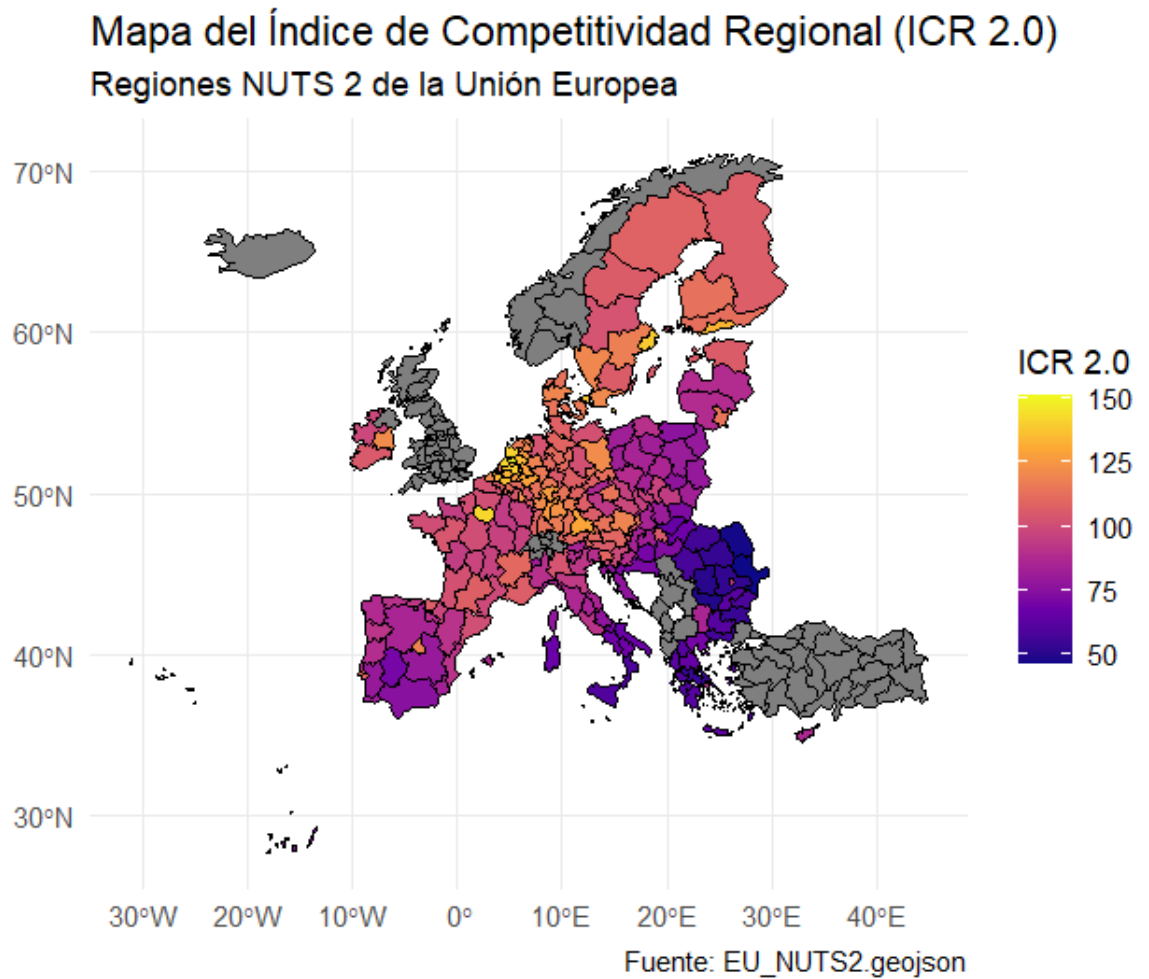
```
# Ver La estructura del nuevo dataset
str(merged_data)

# Verificar Los nombres de Las columnas
names(merged_data)

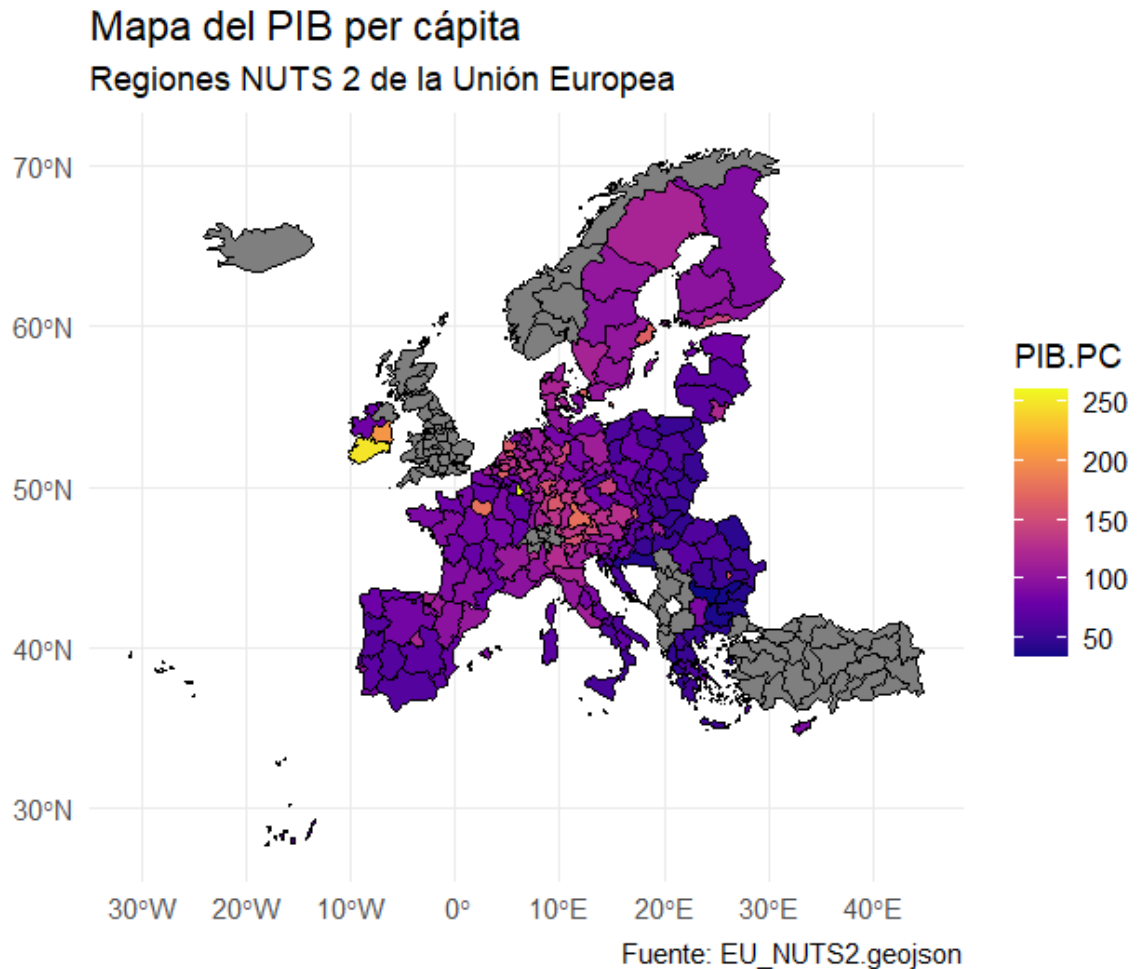
# Resumen general del dataset
summary(merged_data)
```

7.

A continuación se muestra un mapa en el que se exponen los niveles de competitividad regional en una escala de colores:



A continuación, se muestra un mapa en el que se expone el PIB per cápita por región a través de una escala de colores.



Tras observar ambos mapas podemos sospechar de la existencia de una relación entre el PIB per cápita y el índice de competitividad regional.

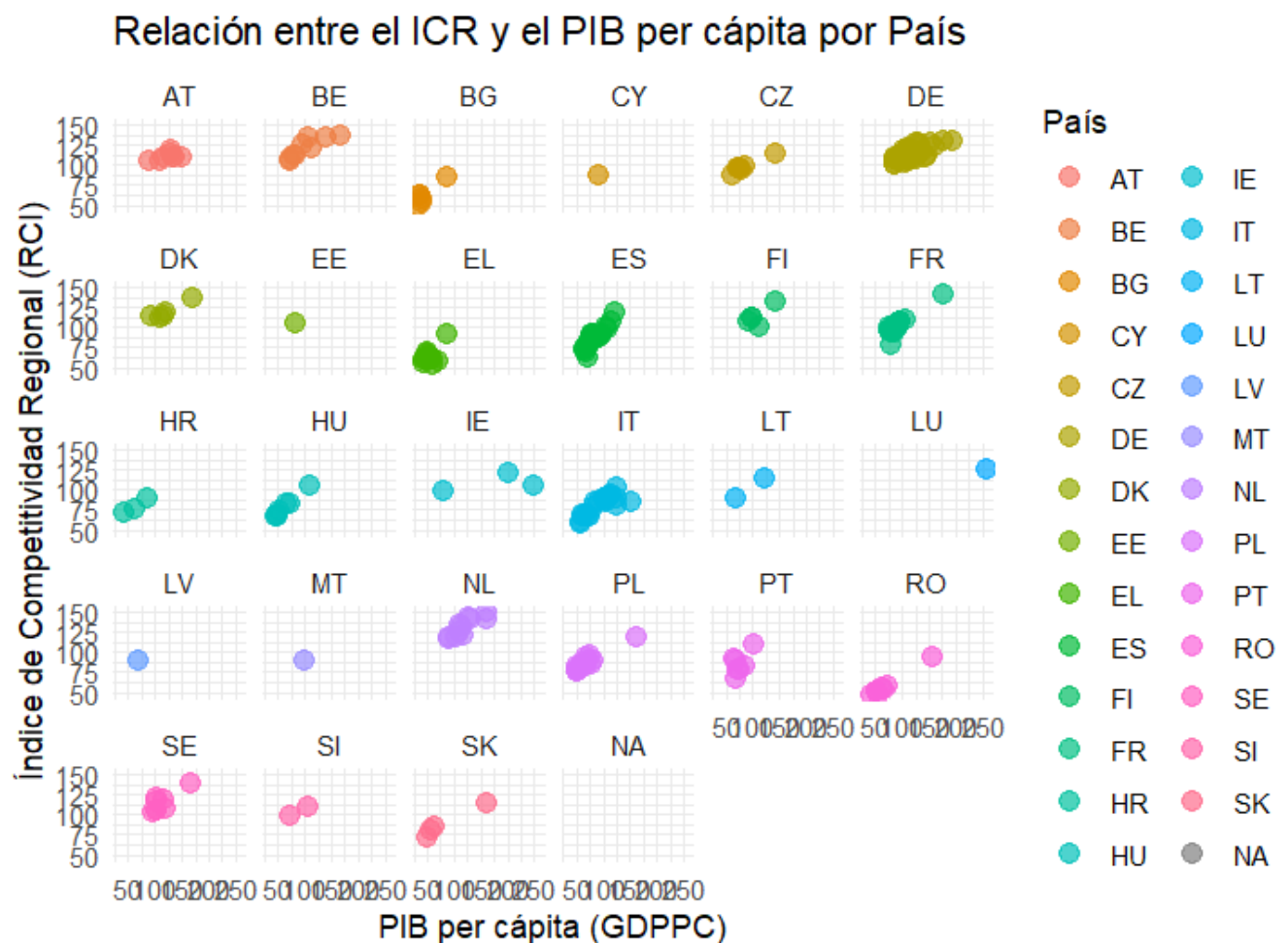
El Código empleado para realizar los respectivos mapas es:

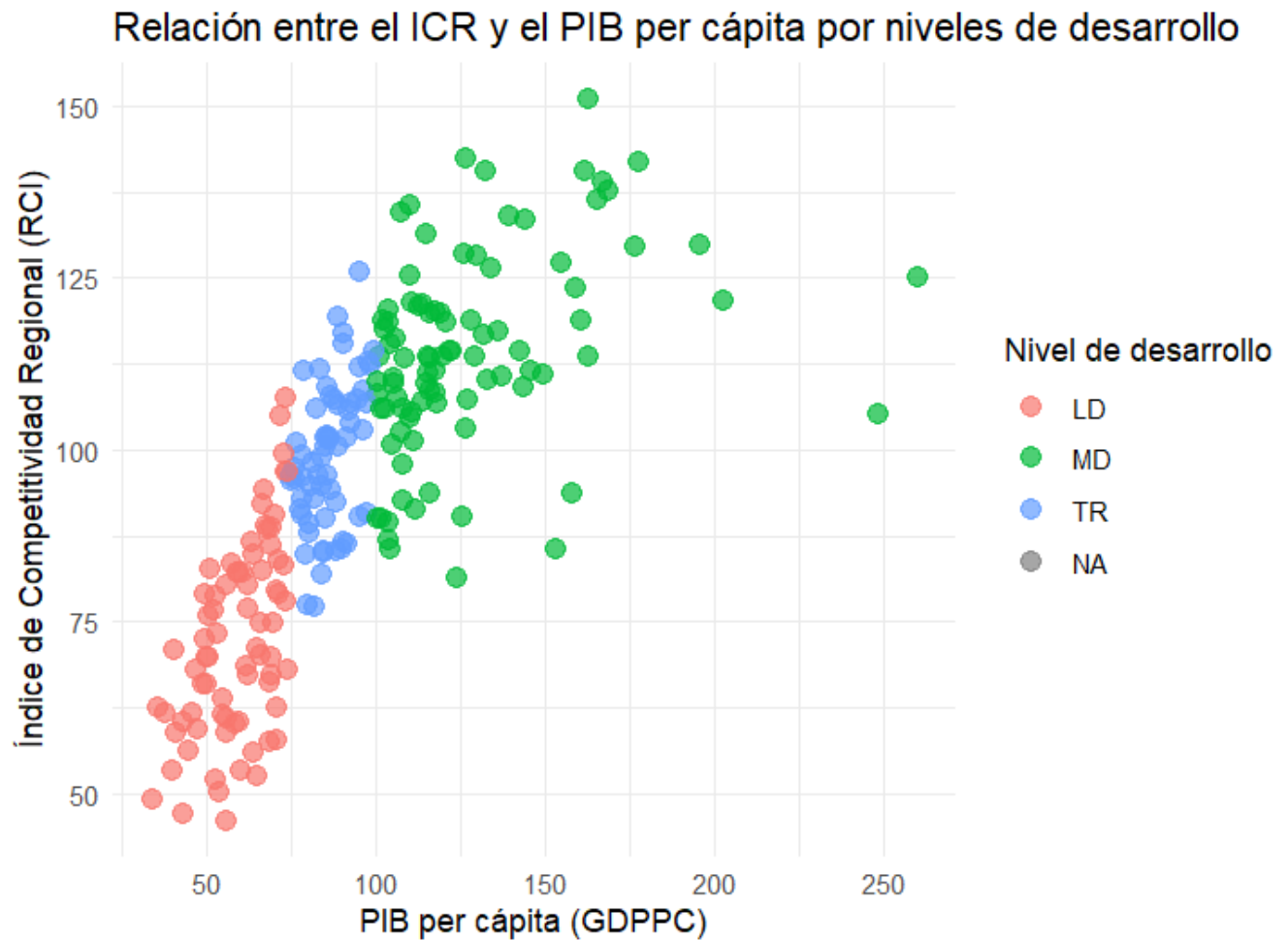
```
ggplot(data = merged_data) +  
  geom_sf(aes(fill = RCI), color = "black", size = 0.1) +  
  scale_fill_viridis_c(option = "plasma", name = "ICR 2.0") +  
  labs(title = "Mapa del Índice de Competitividad Regional (ICR 2.0)",  
        subtitle = "Regiones NUTS 2 de la Unión Europea",  
        caption = "Fuente: EU_NUTS2.geojson") +  
  theme_minimal() +  
  theme(legend.position = "right")
```

```
ggplot(data = merged_data) +
  geom_sf(aes(fill = GDPPC), color = "black", size = 0.1) +
  scale_fill_viridis_c(option = "plasma", name = "PIB.PC") +
  labs(title = "Mapa del PIB per cápita",
       subtitle = "Regiones NUTS 2 de la Unión Europea",
       caption = "Fuente: EU_NUTS2.geojson") +
  theme_minimal() +
  theme(legend.position = "right")
```

8.

En estas dos imágenes se muestra un diagrama de puntos que muestra la relación entre el ICR (eje y) y el PIB per cápita (eje x) segmentando tanto por país como por niveles de desarrollo regional





En este último gráfico podemos observar una relación directa entre el PIB per cápita y el ICR, esta relación tiene un carácter lineal creciente con una pendiente muy pronunciada en el caso de LD (Less developed regions) mientras que en las regiones TR (transition regions) se mantiene la relación lineal directa, con una pendiente más pronunciada. Por último, en las MD (More developed regions) no se observa una relación tan fuerte entre ambas variables, aunque si que es directa.

El código desarrollado para realizar los gráficos anteriores es:

```
ggplot(merged_data, aes(x = GDPPC, y = RCI, color = DEVLEV)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(title = "Relación entre el ICR y el PIB per cápita por niveles
de desarrollo",
       x = "PIB per cápita (GDPPC)",
       y = "Índice de Competitividad Regional (RCI)",
       color = "Nivel de desarrollo") +
  theme_minimal() +
  theme(legend.position = "right")
```

```
ggplot(merged_data, aes(x = GDPPC, y = RCI, color = CNTR)) +
  geom_point(size = 3, alpha = 0.7) +
```

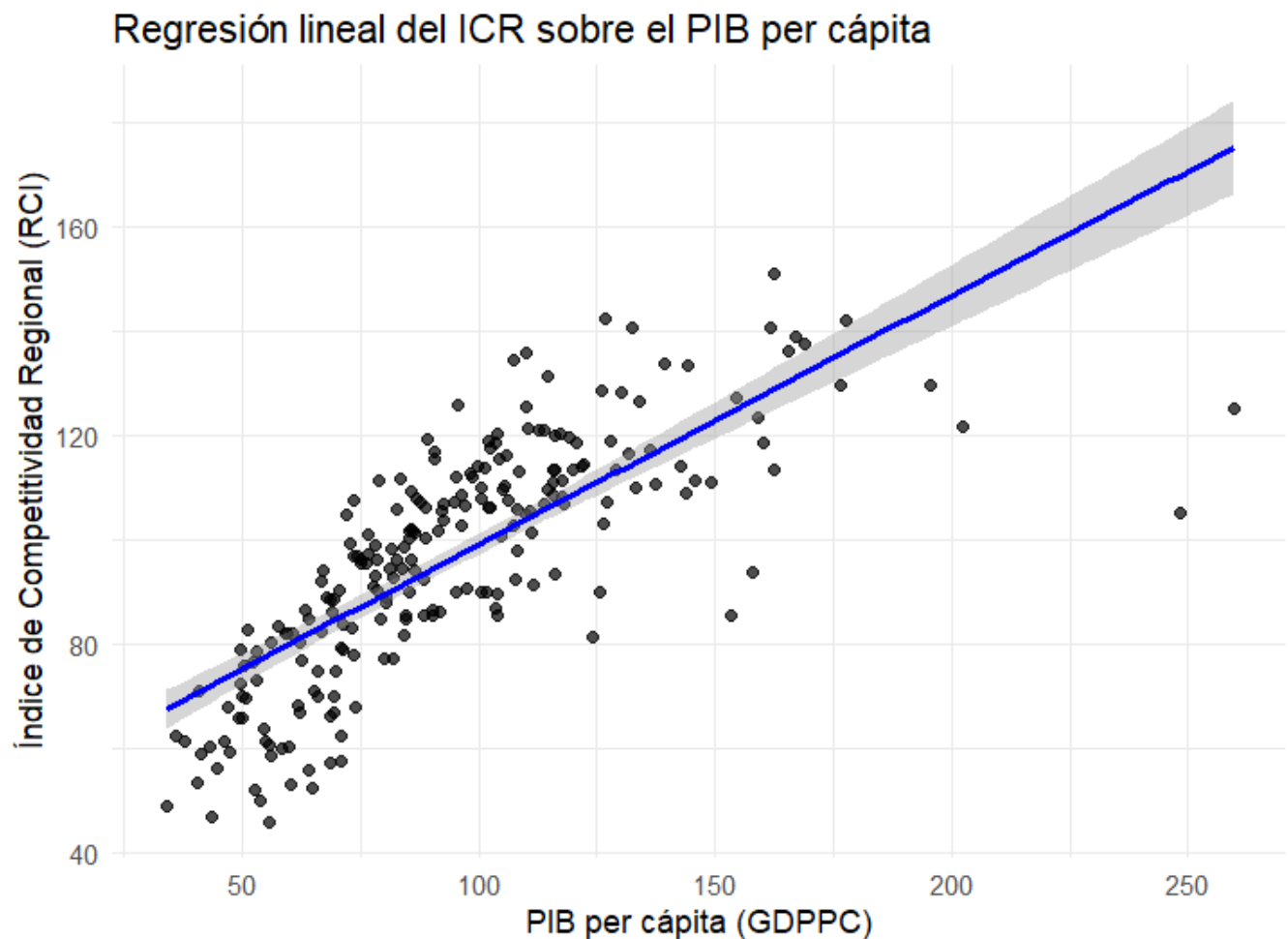
```

facet_wrap(~CNTR) +
labs(title = "Relación entre el ICR y el PIB per cápita por País",
     x = "PIB per cápita (GDPPC)",
     y = "Índice de Competitividad Regional (RCI)",
     color = "País") +
theme_minimal() +
theme(legend.position = "right")

```

9.

A continuación podemos observar un diagrama de puntos que permite estudiar la relación entre el ICR (eje y) y el PIB per cápita (eje x):



Esta recta de regresión presenta un R^2 de 0.5825 y un error residual de 14.17, lo que indica una relación directa de grado medio entre las variables ICR y PIB per cápita, es decir, cuanto mayor sea el PIB per cápita de una región mayor será su ICR y viceversa.

El código implementado para desarrollar el gráfico es el siguiente:

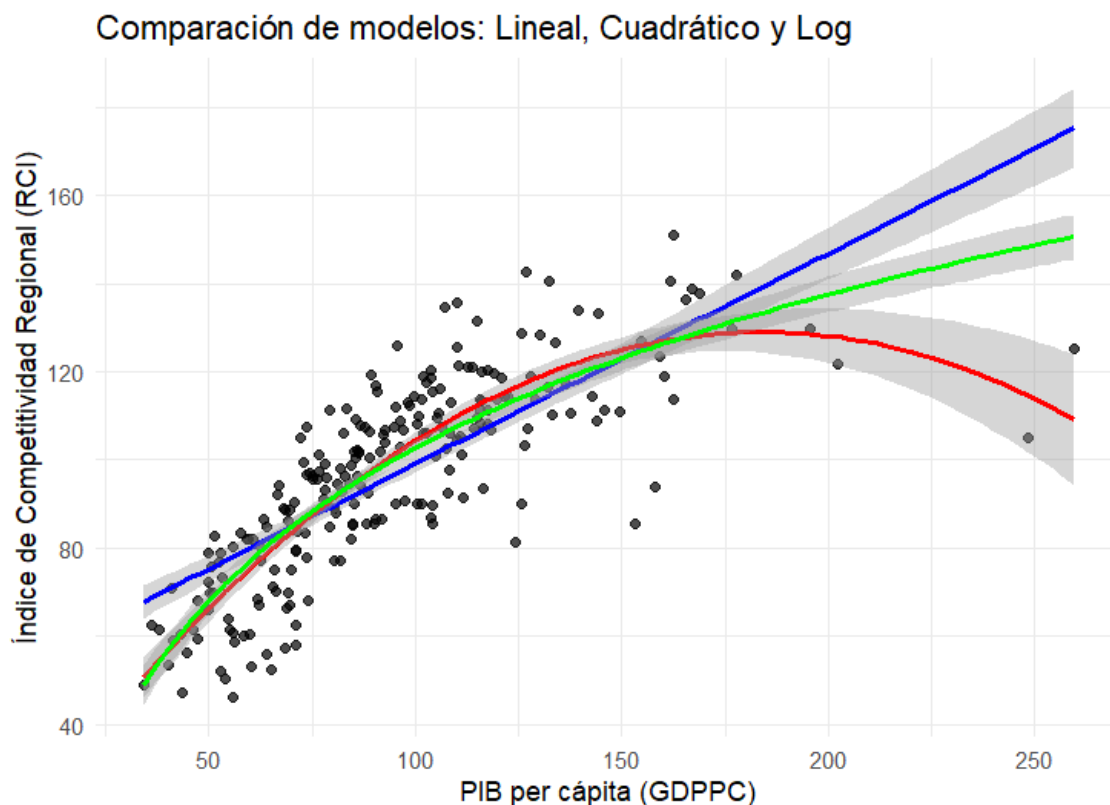
```
# Crear el gráfico de dispersión con la línea de regresión
ggplot(merged_data, aes(x = GDPPC, y = RCI)) +
  geom_point(alpha = 0.7) + # Añadir puntos
  geom_smooth(method = "lm", se = TRUE, color = "blue") + labs(title =
"Regresión lineal del ICR sobre el PIB per cápita",
  x = "PIB per cápita (GDPPC)",
  y = "Índice de Competitividad Regional (RCI)") +
  theme_minimal()

modelo_lineal <- lm(RCI ~ GDPPC, data = merged_data)

summary(modelo_lineal)
```

10.

Una comparación gráfica de los 3 modelos:



En el caso de la regresión cuadrática obtenemos un R^2 de 0.7121 y un error residual de 12.05, es decir, este modelo se ajusta mejor a los datos que el modelo lineal.

Por otra parte, el modelo logarítmico presenta un R^2 de 0.6835 y un error residual de 12.6, es decir, es ligeramente peor que el modelo anterior.

Estudiando los AIC de los 3 modelos observamos lo siguiente:

##		df	AIC
##	modelo_regresion	3	1877.8158
##	modelo_cuadratico	4	1794.6888
##	modelo_log	3	1814.4125

Como sospechábamos el modelo con menor AIC es el modelo cuadrático, por tanto siguiendo el criterio de AIC el modelo que mejor se ajusta a los datos es el modelo cuadrático

El código implementado para desarrollar el gráfico es el siguiente:

```
# Ajustar el modelo lineal
modelo_lineal <- lm(RCI ~ GDPPC, data = merged_data)

# Mostrar el resumen del modelo lineal
summary(modelo_lineal)

# Crear el término cuadrático del PIB per cápita
modelo_cuadratico <- lm(RCI ~ GDPPC + I(GDPPC^2), data = merged_data)

# Mostrar el resumen del modelo cuadrático
summary(modelo_cuadratico)

# Ajustar el modelo Log (Logarítmico)
modelo_log <- lm(RCI ~ log(GDPPC), data = merged_data)

# Mostrar el resumen del modelo log
summary(modelo_log)

# Comparar AIC entre los modelos
AIC(modelo_regresion, modelo_cuadratico, modelo_log)

ggplot(merged_data, aes(x = GDPPC, y = RCI)) +
  geom_point(alpha = 0.7) + # Puntos de datos
  # Línea de regresión lineal
  geom_smooth(method = "lm", se = TRUE, color = "blue", formula = y ~
x) +
  # Línea de regresión cuadrática
  geom_smooth(method = "lm", se = TRUE, color = "red", formula = y ~ p
oly(x, 2)) +
  # Línea de regresión log-log (con inversión logarítmica de los valor
es)
  geom_smooth(method = "lm", se = TRUE, color = "green", formula = y ~
log(x)) +
  labs(title = "Comparación de modelos: Lineal, Cuadrático y Log-log",
    x = "PIB per cápita (GDPPC)",
    y = "Índice de Competitividad Regional (RCI)") +
  theme_minimal()
```