

# Az internethasználat és a női munkaerő-piaci részvétel elemzése

Csoport [SZÁM] - [TAG 1], [TAG 2], [TAG 3], [TAG 4]

2025. november 4.

## Bevezetés

Ez a dokumentum elemzi az internethasználat aránya és a női munkaerő-piaci részvételi arány közötti kapcsolatot különböző országokban 2022-re vonatkozóan. Az adatok a Világbanktól származnak.

## Kutatási kérdés és hipotézis

**Kutatási kérdés:** Van-e kapcsolat az internethasználat és a női munkaerő-piaci részvételi arány között az országok tekintetében?

**Hipotézis:** **Pozitív kapcsolatot** várunk az internethasználat és a női munkaerő-piaci részvétel között. Érvelésünk:

- Lehetőségekhez való hozzáférés:** Az internet-hozzáférés jobb hozzáférést biztosít a nők számára az álláslehetőségekhez, beleértve a helyben esetleg nem elérhető távmunka pozíciókat.
- Rugalmasság:** Az online platformok és digitális eszközök rugalmas munkavégzési lehetőségeket tesznek lehetővé, amelyek segíthetnek a nőknek a munka és a családi kötelezettségek hatékonyabb egyensúlyban tartásában.
- Gazdasági fejlődés:** A magasabb internethasználat jellemzően fejlettebb gazdaságokhoz kapcsolódik, amelyekben általában inkább nemek szempontjából egyenlőbb munkaerőpiacok és jobb infrastruktúra támogatja a nők munkaerő-piaci részvételét.
- Oktatás és készségek:** Az internet-hozzáférés elősegíti az online oktatást és készségfejlesztést, potenciálisan növelve a nők foglalkoztathatóságát.

Ugyanakkor elismerjük, hogy ez a kapcsolat összetett lehet, és számos kulturális, gazdasági és szociális tényező befolyásolhatja, amelyek országonként változnak.

## 1. Adatok előkészítése

Először betöljük a szükséges könyvtárakat. A `tidyverse` csomagot fogjuk használni adatkezeléshez és vizualizációhoz, valamint a `WDI` csomagot az adatok közvetlen letöltéséhez a Világbanktól.

```
library(tidyverse)
library(WDI)
```

Most letöljük az adatokat közvetlenül a Világbanktól a `WDI` csomag használatával. Két mutatóra van szükségünk: - IT.NET.USER.ZS: Internethasználat (a népesség %-ában) - SL.TLF.CACT.FE.ZS: Női munkaerő-piaci részvételi arány (%)

```

# Definiáljuk a letölteni kívánt mutatókat
indicators <- c("IT.NET.USER.ZS", "SL.TLF.CACT.FE.ZS")

# Letöltsük az adatokat a Világbanktól a 2022-es évre
wb_data <- WDI(
  country = "all",
  indicator = indicators,
  start = 2022,
  end = 2022,
  extra = TRUE
)

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Warning in open.connection(con, "rb"): cannot open URL
## 'https://api.worldbank.org/v2/en/country/all/indicator/SL.TLF.CACT.FE.ZS?format=json&date=2022:2022&'
## HTTP status was '429 Unknown Error'

## Előkészítjük az adatokat
final_data <- wb_data %>%
  select(country, IT.NET.USER.ZS, SL.TLF.CACT.FE.ZS) %>%
  rename(
    Country.Name = country,
    internet_usage = IT.NET.USER.ZS,
    female_labor_force = SL.TLF.CACT.FE.ZS

```

```

) %>%
na.omit()

# Megjelenítjük a végeleges adathalmaz első néhány sorát
head(final_data)

##          Country.Name internet_usage female_labor_force
## 1      Afghanistan     17.1917           5.15900
## 2 Africa Eastern and Southern    30.2000          60.44514
## 3   Africa Western and Central   36.9000          64.73852
## 4            Albania        82.6137          53.18300
## 5            Algeria       74.8319          14.09200
## 8            Angola        42.0719          73.17900

```

## 2. Leíró statisztikák

Itt kiszámítjuk az alapvető leíró statisztikákat a két változónkra. Ez tartalmazza az átlagot, mediánt, minimumot, maximumot, kvartiliseket, varianciát, tartományt és az interkvartilis tartományt.

```

# Készítünk egy szép összefoglaló táblázatot
desc_stats <- data.frame(
  Statistic = c("Átlag", "Medián", "Szórás", "Variancia", "Minimum", "Maximum", "Tartomány", "IQR"),
  Internet_Usage = c(
    round(mean(final_data$internet_usage), 2),
    round(median(final_data$internet_usage), 2),
    round(sd(final_data$internet_usage), 2),
    round(var(final_data$internet_usage), 2),
    round(min(final_data$internet_usage), 2),
    round(max(final_data$internet_usage), 2),
    round(diff(range(final_data$internet_usage)), 2),
    round(IQR(final_data$internet_usage), 2)
  ),
  Female_Labor_Force = c(
    round(mean(final_data$female_labor_force), 2),
    round(median(final_data$female_labor_force), 2),
    round(sd(final_data$female_labor_force), 2),
    round(var(final_data$female_labor_force), 2),
    round(min(final_data$female_labor_force), 2),
    round(max(final_data$female_labor_force), 2),
    round(diff(range(final_data$female_labor_force)), 2),
    round(IQR(final_data$female_labor_force), 2)
  )
)

knitr::kable(desc_stats,
  col.names = c("Statisztika", "Internethasználat (%)", "Női munkaerő-piaci részvétel (%)" ),
  caption = "Mindkét változó leíró statisztikái",
  align = c("l", "r", "r")
)

```

Table 1: Mindkét változó leíró statisztikái

Statisztika	Internethasználat (%)	Női munkaerő-piaci részvétel (%)
Átlag	67.73	50.57
Medián	76.50	52.56
Szórás	25.15	14.51
Variancia	632.35	210.52
Minimum	11.00	5.16
Maximum	100.00	82.58
Tartomány	89.00	77.42
IQR	40.36	16.20

**Értelmezés:** Az országok átlagos internethasználata 67,7%, nagy változékonysággal (89%-os tartomány). A női munkaerő-piaci részvétel átlaga 50,6%, még nagyobb tartománnyal (77,4%), ami jelentős globális egyenlőtlenségeket jelez a nők munkaerő-piaci részvételében.

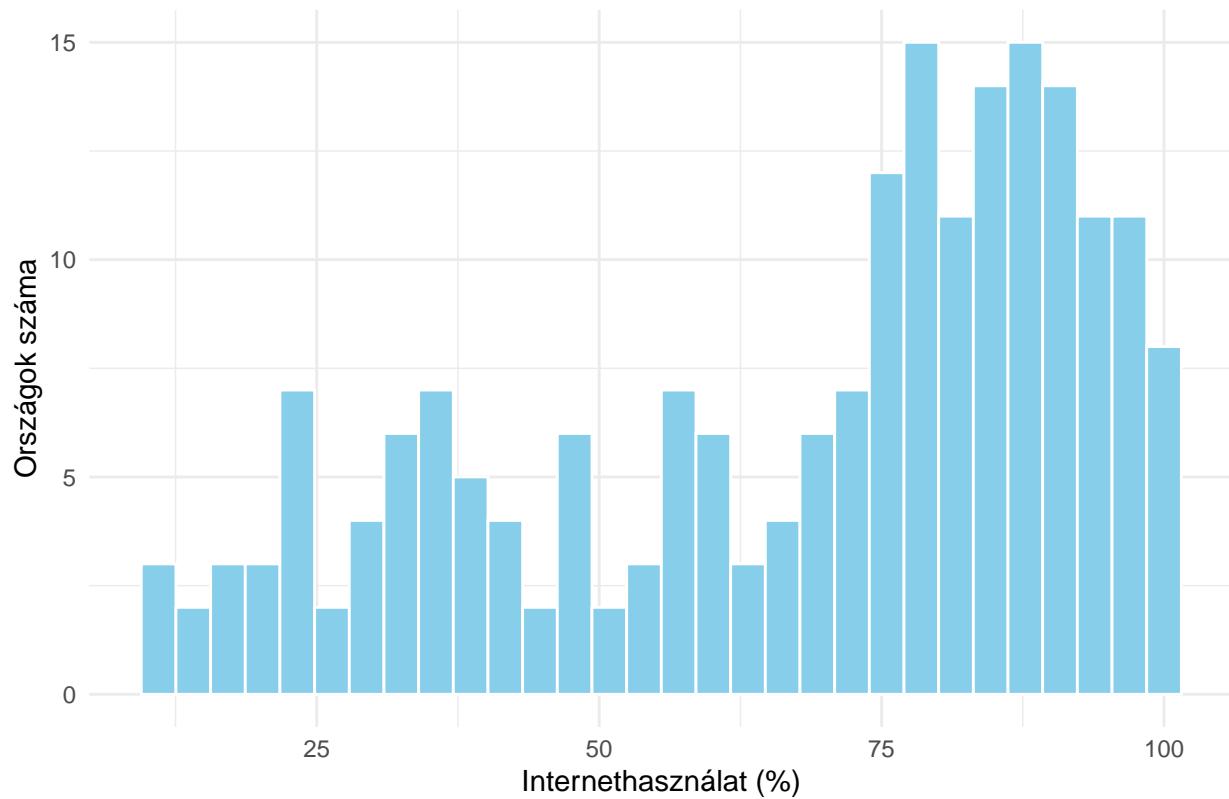
### 3. Adatvizualizáció

Az adatok vizualizálása segít megérteni minden változó eloszlását és azonosítani a potenciális kiugró értékeket.

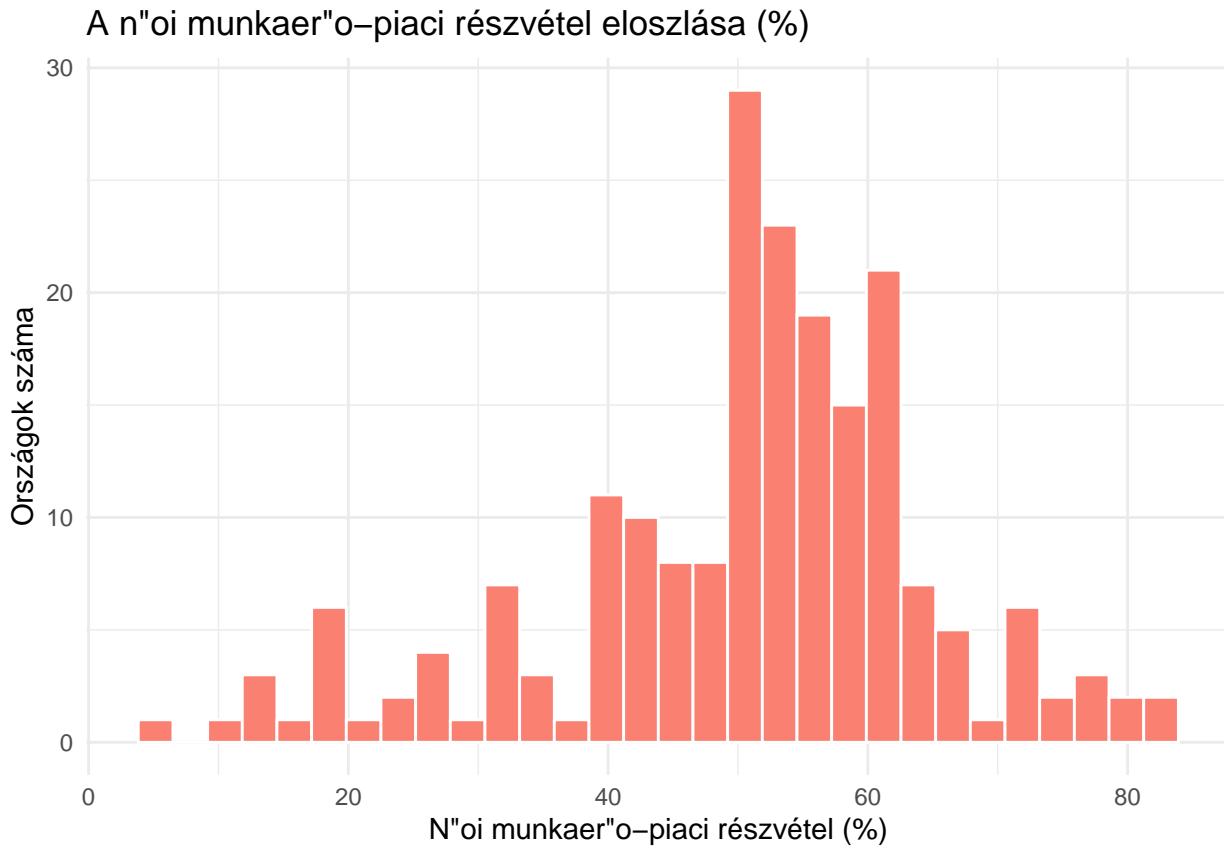
**Hisztogramok** A hisztogramok egy változó eloszlását mutatják.

```
# Hisztogram az internethasználatra
ggplot(final_data, aes(x = internet_usage)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "white") +
  labs(
    title = "Az internethasználat eloszlása (%)",
    x = "Internethasználat (%)",
    y = "Országok száma"
  ) +
  theme_minimal()
```

Az internethasználat eloszlása (%)



```
# Hisztogram a női munkaerő-piaci részvételre
ggplot(final_data, aes(x = female_labor_force)) +
  geom_histogram(bins = 30, fill = "salmon", color = "white") +
  labs(
    title = "A női munkaerő-piaci részvétel eloszlása (%)",
    x = "Női munkaerő-piaci részvétel (%)",
    y = "Országok száma"
  ) +
  theme_minimal()
```

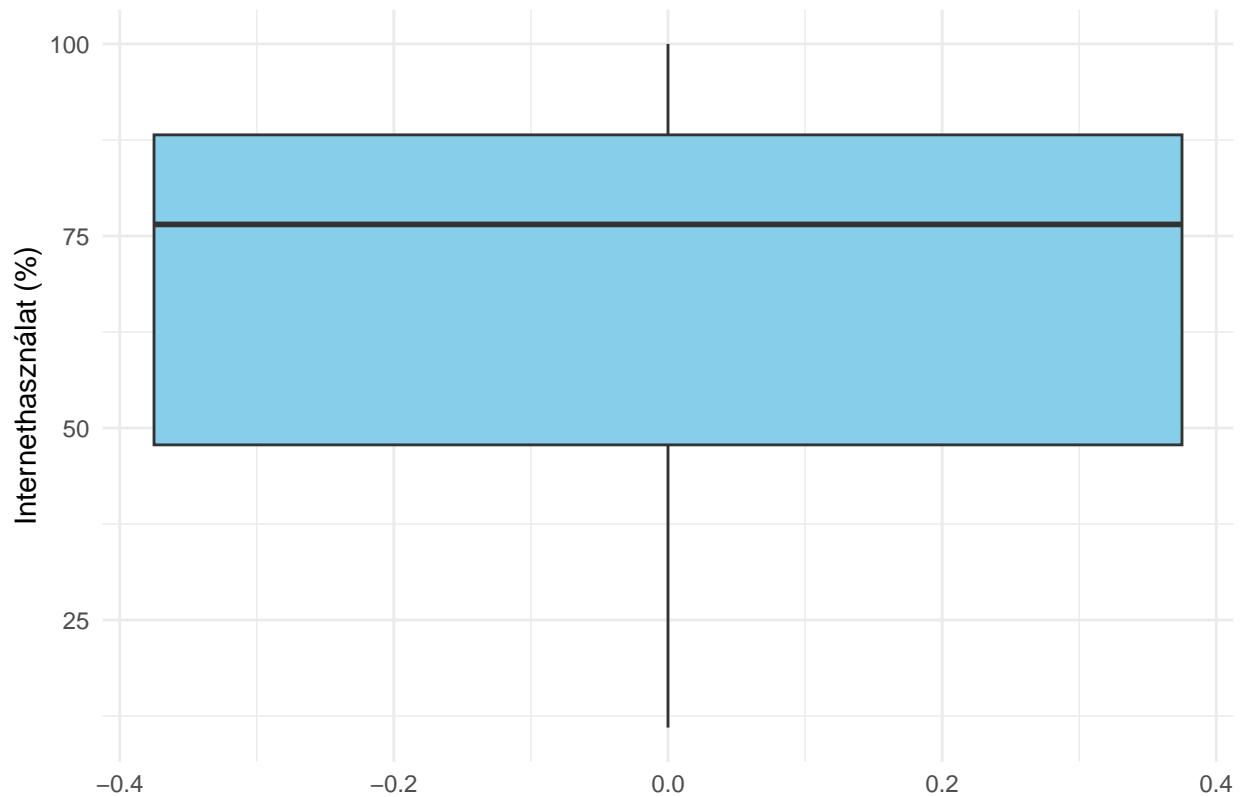


**Értelmezés:** Az internethasználati hisztogram balra ferde eloszlást mutat, a legtöbb ország magas internet-penetrációval (70-90%) rendelkezik, míg kevesebb országban alacsony a használat. A női munkaerő-piaci részvételi hisztogram normálisabban eloszló, enyhe balra ferdeséggel, 50% körül középpontosulva.

**Dobozdiagramok** A dobozdiagramok hasznosak a kiugró értékek azonosításához és az adatok eloszlásának megtekintéséhez.

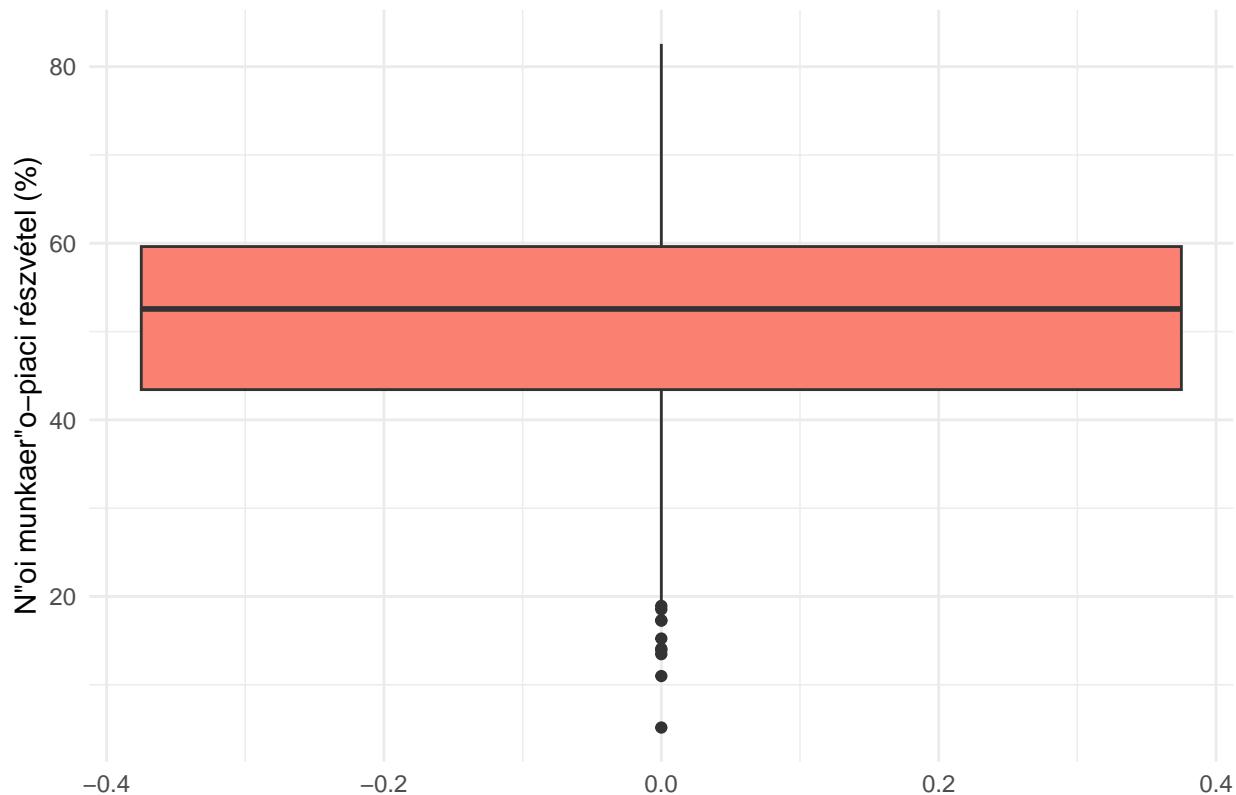
```
# Dobozdiagram az internethasználatra
ggplot(final_data, aes(y = internet_usage)) +
  geom_boxplot(fill = "skyblue") +
  labs(
    title = "Az internethasználat dobozdiagramja (%)",
    y = "Internethasználat (%)"
  ) +
  theme_minimal()
```

Az internethasználat dobozdiagramja (%)



```
# Dobozdiagram a női munkaerő-piaci részvételre
ggplot(final_data, aes(y = female_labor_force)) +
  geom_boxplot(fill = "salmon") +
  labs(
    title = "A női munkaerő-piaci részvétel dobozdiagramja (%)",
    y = "Női munkaerő-piaci részvétel (%)"
  ) +
  theme_minimal()
```

A n"oi munkaer"o-piaci részvétel dobozdiagramja (%)



**Értelmezés:** A dobozdiagramok számos kiugró értéket tárnak fel. A női munkaerő-piaci részvételnél vannak extrém alacsony értékekkel (20% alatt) rendelkező országok, amelyek olyan országokat képviselhetnek, ahol jelentős kulturális vagy jogi akadályok vannak a nők munkaerő-piaci részvételével szemben.

#### 4. Extrém értékek elemzése

Az IQR módszer használatával azonosítjuk és megvizsgáljuk az adathalmazunk extrém értékeit (kiugró értékeket).

```
# Függvény a kiugró értékek azonosítására az IQR módszerrel
detect_outliers <- function(x) {
  q1 <- quantile(x, 0.25)
  q3 <- quantile(x, 0.75)
  iqr <- q3 - q1
  lower_bound <- q1 - 1.5 * iqr
  upper_bound <- q3 + 1.5 * iqr
  return(x < lower_bound | x > upper_bound)
}

# Kiugró értékek azonosítása az internethasználtnál
internet_outliers <- final_data[detect_outliers(final_data$internet_usage), ]
if (nrow(internet_outliers) > 0) {
  internet_outliers_sorted <- internet_outliers[
    order(internet_outliers$internet_usage),
    c("Country.Name", "internet_usage", "female_labor_force")]
}
```

```

knitr::kable(internet_outliers_sorted,
  col.names = c("Country", "Internet Usage (%)", "Female Labor Force (%)" ),
  caption = "Countries with Extreme Internet Usage Values",
  row.names = FALSE,
  align = c("l", "r", "r")
)
}

# Detect outliers for female labor force
labor_outliers <- final_data[detect_outliers(final_data$female_labor_force), ]
if (nrow(labor_outliers) > 0) {
  labor_outliers_sorted <- labor_outliers[
    order(labor_outliers$female_labor_force),
    c("Country.Name", "internet_usage", "female_labor_force")]
}

knitr::kable(labor_outliers_sorted,
  col.names = c("Country", "Internet Usage (%)", "Female Labor Force (%)" ),
  caption = "Countries with Extreme Female Labor Force Participation Values",
  row.names = FALSE,
  align = c("l", "r", "r"))
}
}

```

Table 2: Countries with Extreme Female Labor Force Participation Values

Country	Internet Usage (%)	Female Labor Force (%)
Afghanistan	17.1917	5.15900
Iraq	78.7156	10.99000
Iran, Islamic Rep.	79.0610	13.46900
Jordan	90.5000	13.92200
Algeria	74.8319	14.09200
Egypt, Arab Rep.	72.1981	15.23300
Middle East, North Africa, Afghanistan & Pakistan (IDA & IBRD)	52.0000	17.27636
Middle East, North Africa, Afghanistan & Pakistan (excluding high income)	52.3000	17.28747
Djibouti	63.2573	18.54600
West Bank and Gaza	88.6469	18.93400

**Interpretation:** The extreme values analysis reveals countries that deviate significantly from the typical patterns. Countries with extremely low female labor force participation (below 20%) often represent nations with specific cultural, religious, or economic factors that limit women's workforce participation. These outliers are important to identify as they may have disproportionate influence on our correlation analysis.

## 5. Correlation Analysis

Now, we'll quantify the linear relationship between our two variables by calculating the Pearson correlation coefficient. The coefficient ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation). A value near 0 suggests no linear correlation.

```

# Korrelációs teszt elvégzése
cor_test <- cor.test(final_data$internet_usage, final_data$female_labor_force)

# Szép táblázat készítése a korrelációs eredményekhez
cor_results <- data.frame(
  Metric = c("Korrelációs együttható", "t-statisztika", "p-érték", "95% CI alsó", "95% CI felső"),
  Value = c(
    round(cor_test$estimate, 4),
    round(cor_test$statistic, 4),
    round(cor_test$p.value, 4),
    round(cor_test$conf.int[1], 4),
    round(cor_test$conf.int[2], 4)
  )
)

knitr::kable(cor_results,
  col.names = c("Mutató", "Érték"),
  caption = "Pearson korrelációs teszt eredményei",
  align = c("l", "r"))
)

```

Table 3: Pearson korrelációs teszt eredményei

Mutató	Érték
Korrelációs együttható	-0.0597
t-statisztika	-0.8483
p-érték	0.3973
95% CI alsó	-0.1958
95% CI felső	0.0786

**Értelmezés:** A -0,06-os korrelációs együttható nagyon gyenge negatív kapcsolatot jelez az internethasználat és a női munkaerő-piaci részvétel között. Ez az eredmény **ellentétes a kezdeti hipotézisünkkel**, amely pozitív kapcsolatot jóolt. A korreláció statisztikailag nem szignifikáns ( $p\text{-érték} > 0,05$ ), ami arra utal, hogy országszinten az internethasználat önmagában nem jósolja meg a női munkaerő-piaci részvételt. Ez a megállokáció rávilágít arra, hogy a kulturális, gazdasági és szakpolitikai tényezők fontosabb szerepet játszhatnak, mint a technológiai hozzáférés.

## 6. Pontdiagram és vizuális elemzés

A pontdiagram a legjobb módja két folytonos változó közötti kapcsolat vizualizálásának. Egy regressziós vonalat is hozzáadunk, hogy tisztábban lássuk a trendet.

```

ggplot(final_data, aes(x = internet_usage, y = female_labor_force)) +
  geom_point(alpha = 0.6, color = "steelblue", size = 2) +
  geom_smooth(method = "lm", col = "red", se = TRUE) + # "lm" jelentése: lineáris modell
  labs(
    title = "Internethasználat vs. női munkaerő-piaci részvétel (2022)",
    x = "Internethasználat (a népesség %-ában)",
    y = "Női munkaerő-piaci részvétel (%)",
    caption = "Forrás: Világbank Világfejlesztési Mutatók"
  )

```

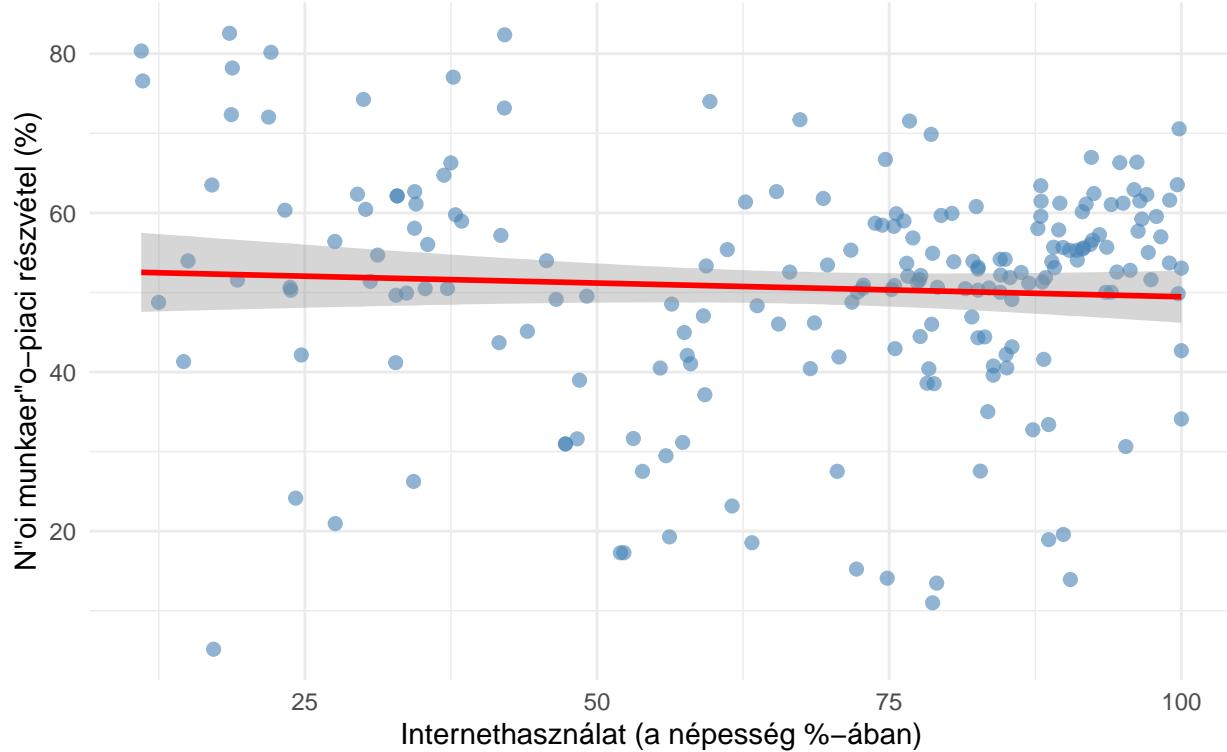
```

theme_minimal() +
theme(plot.title = element_text(face = "bold", size = 14))

```

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

## Internethasználat vs. női munkaerő-piaci részvétel (2022)



Forrás: Világbank Világfejlesztési Mutatók

**Értelmezés:** A pontdiagram közel lapos kapcsolatot tár fel a két változó között, összhangban a gyenge korrelációs együtthatóval. A vörös regressziós vonal enyhe negatív meredekséggel rendelkezik, de a széles konfidencia-intervallum (szürke sáv) nagy bizonytalanságot jelez. A diagram jelentős variációt mutat a női munkaerő-piaci részvételben még a hasonló internethasználati szintekkel rendelkező országok között is, ami arra utal, hogy az internet-hozzáférésen túli egyéb tényezők fontosabb meghatározói a női munkaerő-piaci részvételének. Láthatunk országosoportokat magas internethasználattal, de változó női részvételi szintekkel (40-80%), ami tovább támasztja alá, hogy a kapcsolat nem egyértelmű.

### Következetetés

Ez az elemzés az internethasználat és a női munkaerő-piaci részvétel közötti kapcsolatot vizsgálta 203 országban 2022-ben. Az adatok számos kulcsfontosságú megállapítást tártak fel:

#### Főbb megállapítások:

- Gyenge negatív korreláció:** A -0,06-os korrelációs együttható nagyon gyenge negatív kapcsolatot jelez az internethasználat és a női munkaerő-piaci részvétel között. Ez arra utal, hogy a váratnakkal ellentétben a magasabb internethasználat nem jár együtt a női munkaerő-piaci részvétel növekedésével országszinten.

**2. Internethasználat:** Az országok átlagos internethasználata 67,7% volt, széles tartományban 11%-tól 100%-ig. A 76,5%-os medián azt sugallja, hogy a legtöbb országban viszonylag magas az internet-penetráció.

**3. Női munkaerő-piaci részvétel:** Az átlagos női munkaerő-piaci részvételi arány 50,6% volt, 5,2%-tól 82,6%-ig terjedően. Ez a széles variáció jelentős globális egyenlőtlenségeket emel ki a nők munkaerő-piaci részvételében.

#### **Korlátok és jövőbeli kutatás:**

- Ez az elemzés csak egy 2022-es pillanatképet rögzít, és nem veszi figyelembe az időbeli trendeket vagy az okozati összefüggéseket.
- Az országszintű aggregálás elfedheti a fontos országon belüli eltéréseket és regionális különbségeket.
- Más zavaró tényezők, mint például a kulturális normák, gazdasági fejlettség, oktatási szintek és a gyermekellátás elérhetősége nem kerültek figyelembe vételekre.
- A jövőbeli kutatás feltáráthatná a longitudinális adatokat annak megértésére, hogy az internet-hozzáférés időbeli változásai hogyan kapcsolódnak a női munkaerő-piaci részvétel változásaihoz, és további társadalmi-gazdasági változókat is tartalmazhatna egy átfogóbb elemzéshez.