Female and Patenting Analysis Report

Huiting Wu

2025-03-07

Abstract

This report analyzes the relationship between female researcher and patenting.

Research Questions:

Analyze whether countries with a higher percentage of female researchers, segmented by sector (non-profit, government, business, higher education), demonstrate greater success in intellectual property (IP) creation.

This study examines whether countries with a higher representation in female.

Introduction

Women are underrepresented in research fields, and understanding the relationship between female representation and patents may help address barriers for females in research. If there's a connection between more female researchers and more patents, it could highlight the importance of inclusivity in driving technological progress. Moreover, promoting female researchers may help to create a more productive and innovative research environment.

Key Findings

The relationship between female researchers and patent applications is not straightforward in aggregate from 2017 to 2023. Simply increasing the percentage of female researchers does not always result in more patents.

The strength of this relationship depends on the environment a researcher works in. Without considering specific sectors, only about 1.5% of the variation in patent activity can be linked to the percentage of female researchers in a country. When we consider sectors in it, it increases to about 5%.

Also, the countries with higher percentages of female researchers have stronger positive associations between female researchers and patenting. These suggest that both gender representation and the research environment play a role in patenting.

About Dataset and Key Variables

- The first dataset: **percentage of female researchers.csv**:

There are 3589 rows, 4 column, 83 missing values in value, and 1 duplicated data point

"`Indicators`": Different sectors.

Head count of female researchers;

Full time equivalent of female researcher;

Head count of female researchers in Business enterprise, government, higher education, and non-profit;

Full time equivalent of female researchers in Business enterprise, government, higher education, and non-profit

"`Country`": Country or area names. There are 136 areas. There are 86 countries using for this study after data merging.

"`Time`": Year; 2017 to 2023

"`Value`": percentage of female researchers

- The second dataset: **Total patent applications 2017 to 2023.csv**:

There are 1183 rows, 3 column, and 172 missing values in total patent applications.

"`Origin`": Country names. There are 169 countries.

"`Time`": Year; 2017 to 2023

"`Total Applications`": Total patent applications.

- The third dataset: **Population.csv**

"`Time`": Year; 2017 to 2023

"`Country`": Country or area names. There are 266 areas.

Methods

Data Cleaning

Countries without any data on the percentage of female researchers were removed to ensure completeness in the analysis.

Data Preparation

Dataset Merging: Three datasets were combined based on country names to create a unified dataset.

New Variable Creation: A new variable, patent applications per capita, was computed by dividing the total patent applications by the country's population.

Data Transformation: A log transformation was applied to patent applications per capita to normalize the distribution and improve model assumptions.

Statistical Analysis

ANOVA: Used to compare patent applications per capita across different sectors of female researcher representation. Also, compare the average of female researchers representation in different sectors.

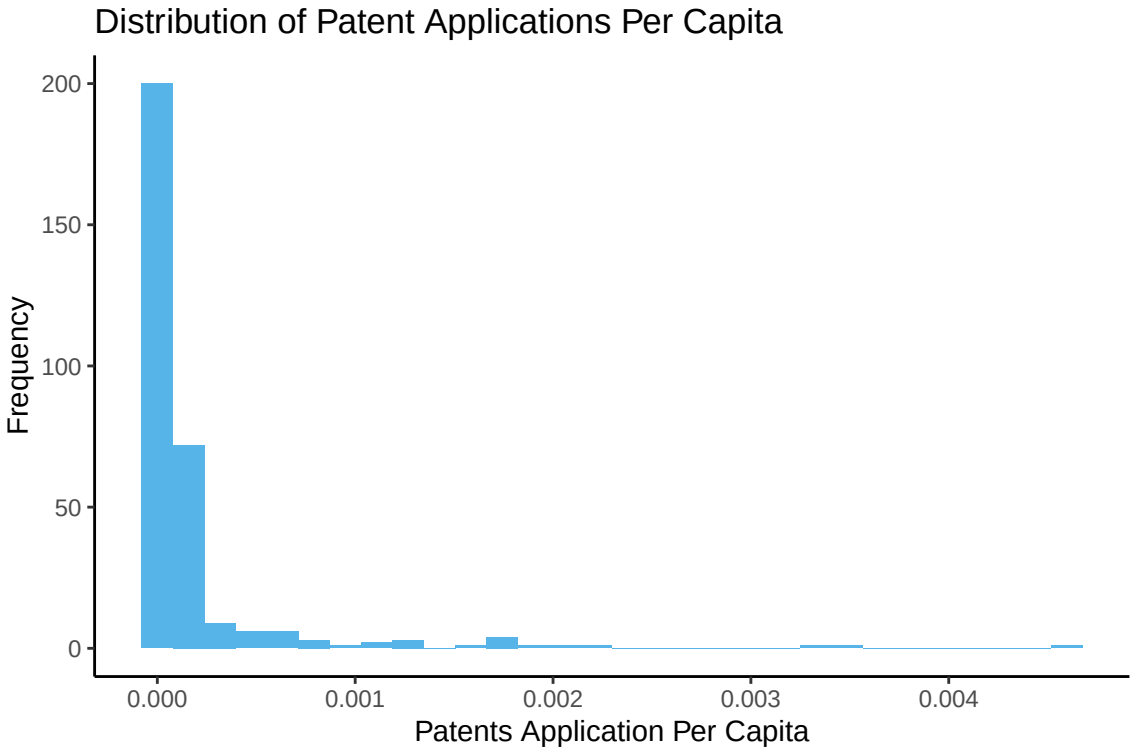
Correlation Test: Measured the strength and direction of the association between female researcher representation and patent applications per capita.

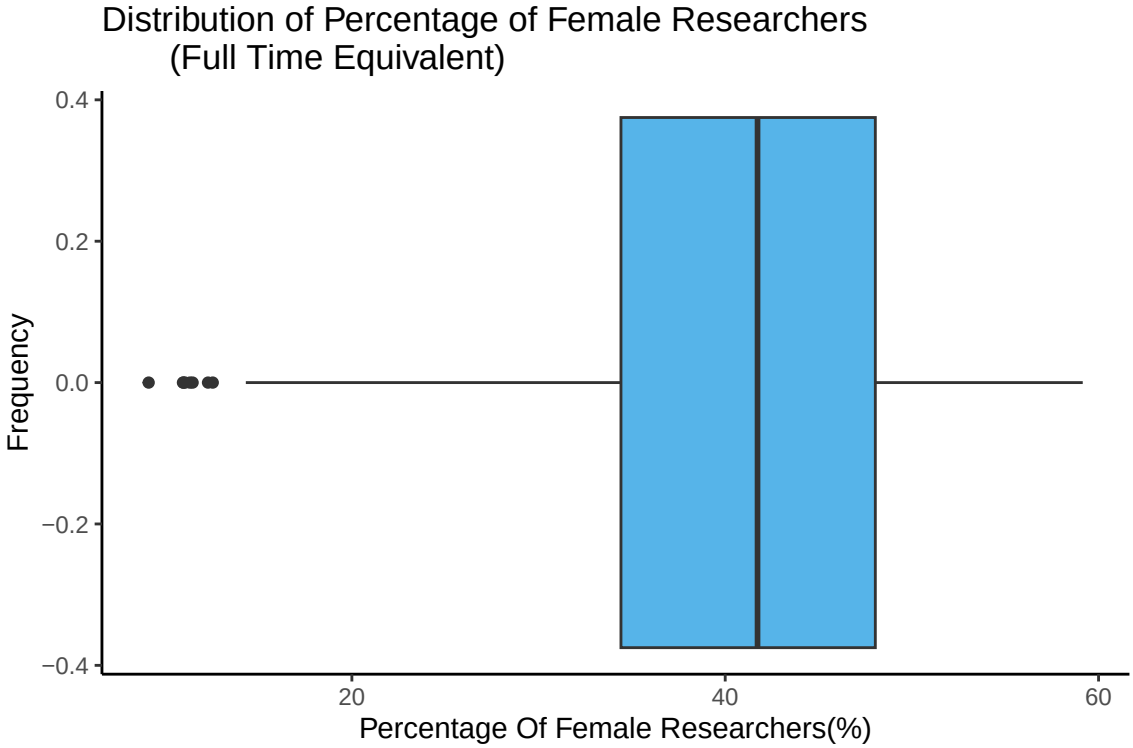
Regression Analysis: Examined the relationship between the percentage of female researchers and patent applications per capita.

ANCOVA: Examined the interaction effect between female researcher representation and research sectors on patent applications per capita.

Data Visualization

Visualize the target variables: Patent applications per capita and percent of female researchers





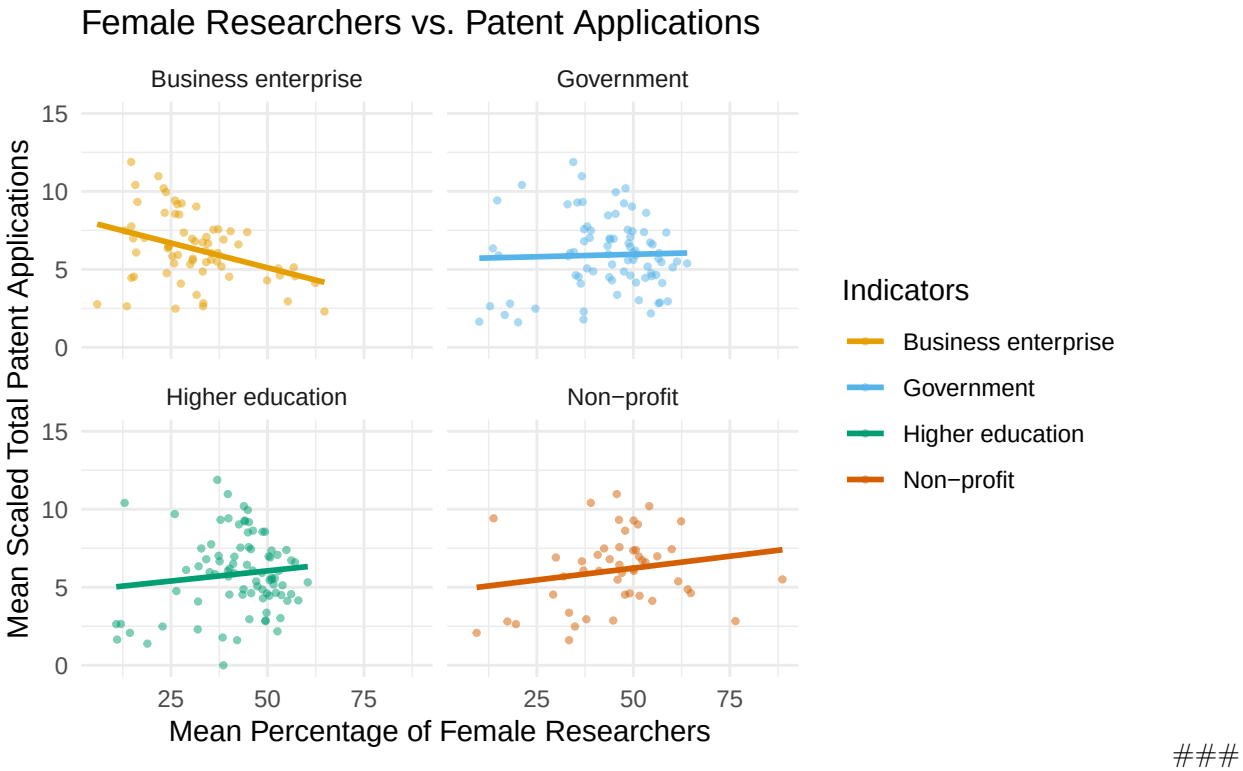
Relationship of percentage of female researchers in different sectors with the mean of log patent applications

Looking at different sectors, there are varied trends. In business, an unexpected trend appears—countries with more female researchers tend to have fewer patents. This could be due to industries like tech and engineering having fewer females or barriers that limit their access to funding and leadership.

In the non-profit sector, some countries have a high percentage of female researchers but low patent activity. This suggests their work may focus more on social impact or policy rather than commercial innovation.

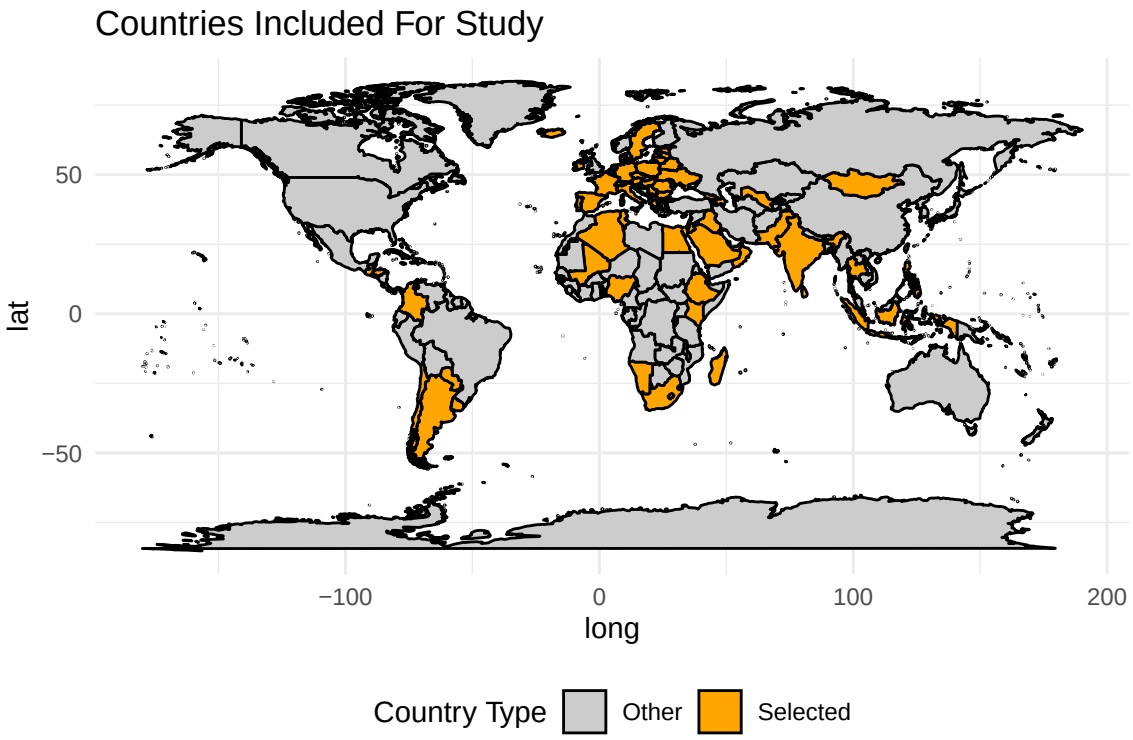
For higher education and government, more female researchers tend to be associated with slightly higher patent applications. This suggests their contributions support innovation, but the effect is not very strong.

Overall, having more female researchers alone isn't enough to boost patent activity. Other factors, like funding, industry focus, and opportunities, play a significant role.



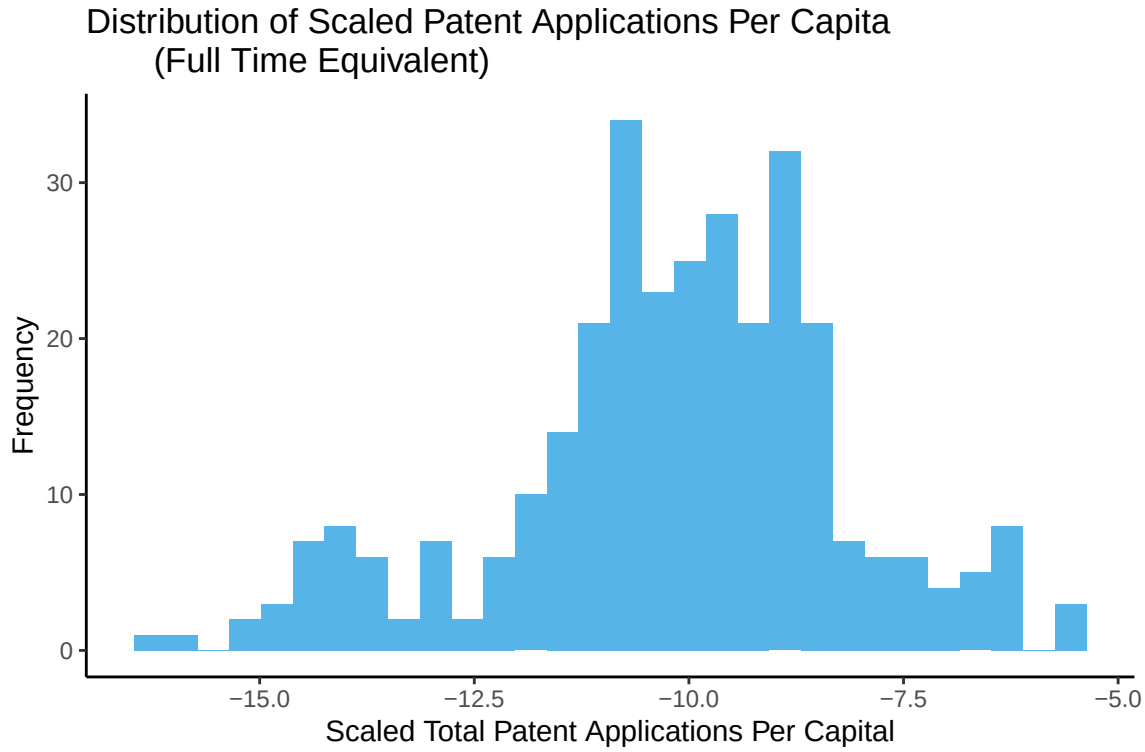
###

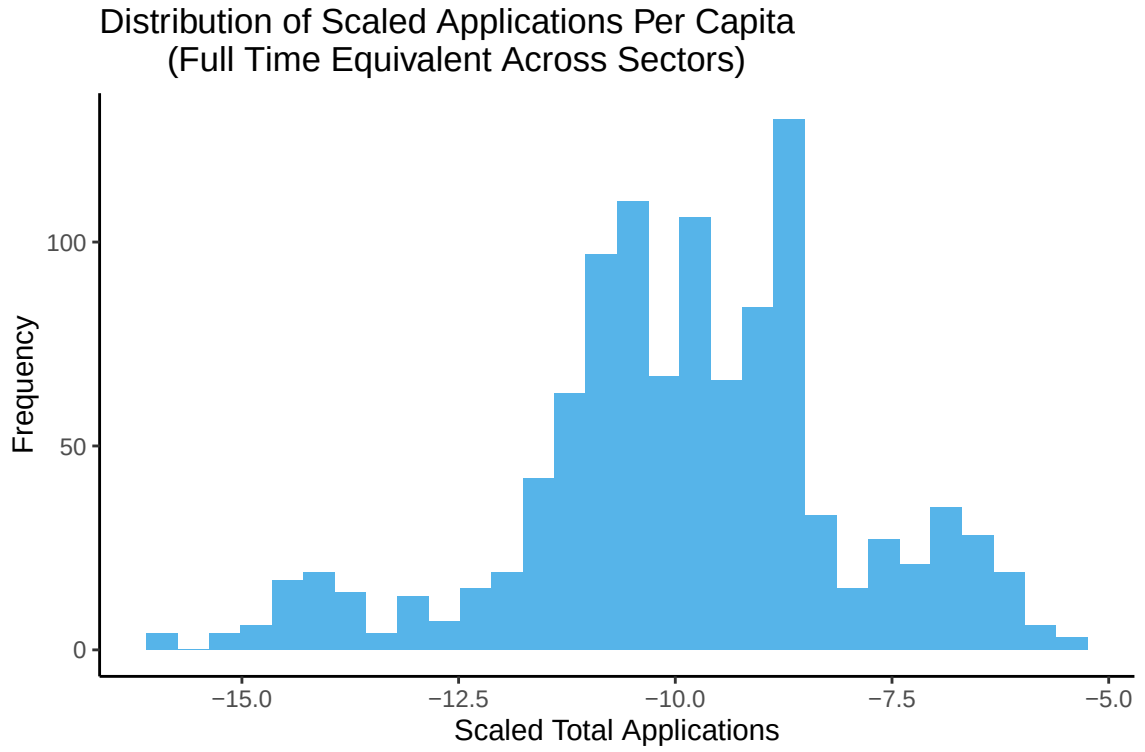
The Location of Countries for the Study



Scaling the dataset

Total applications of patent is in counting scale which is extremely right skewed. Therefore, log transformation is applied.





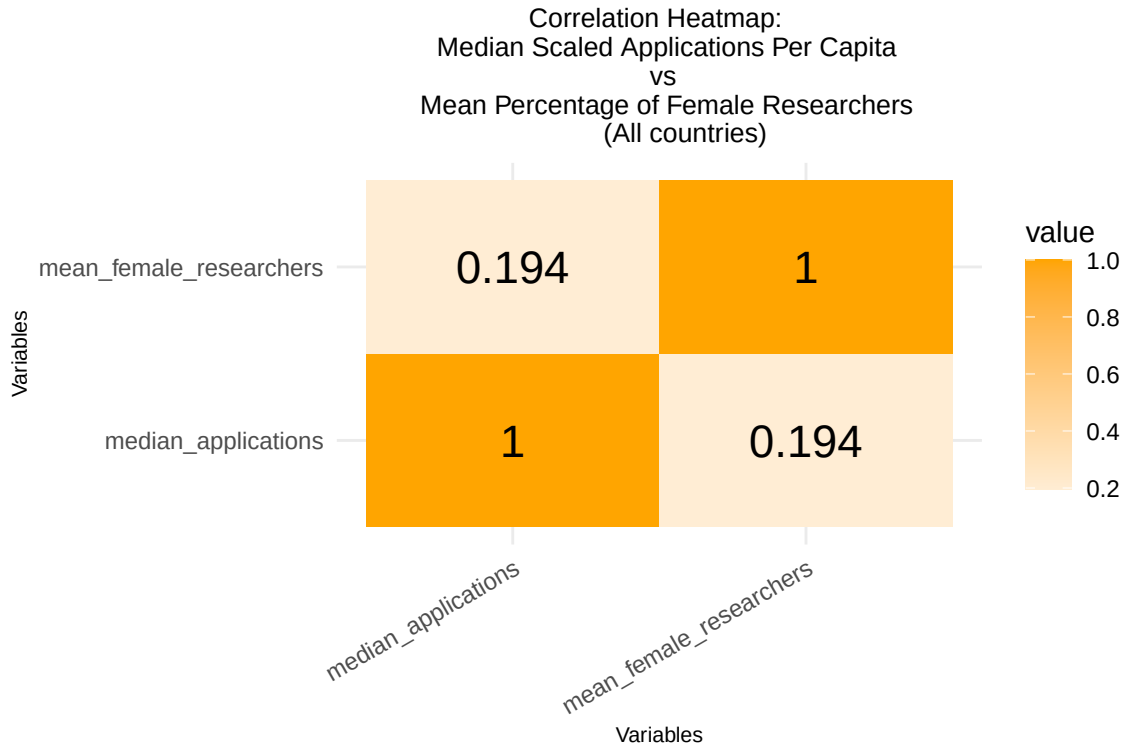
Correlation Coefficient Heat map

These plots shows the linear relationship between female researchers and patent applications.

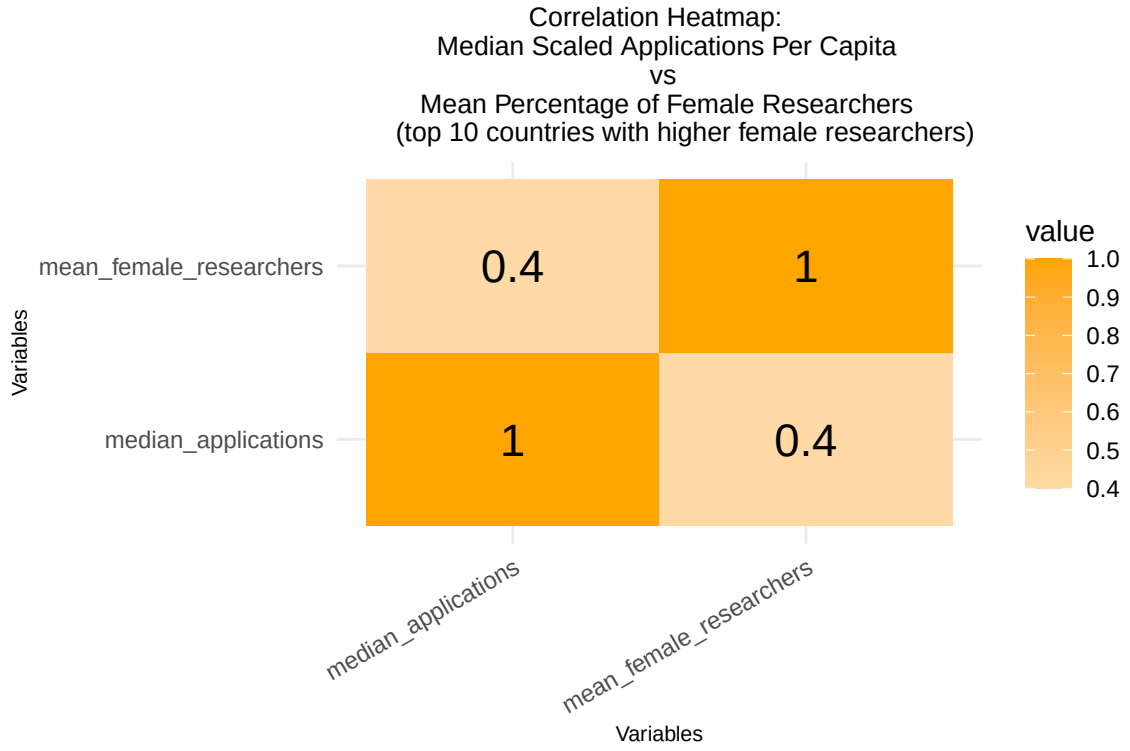
The first plot includes all countries, there is a weak correlation. This means that the relationship is not simply increasing female researchers would result in more patents. Other factors likely influencing the relationship or there are other kinds of relationship between female researchers and patenting like quadratic.

The second plot includes the top 10 countries with a high percentage of female researchers, there is a stronger positive linear relationship. This suggests that in countries where more females are in research, they may have better support, leading to more patents. Stronger policies, funding, and inclusive environments could be key reasons for this difference.

Explore the relationship between the percentage of female researchers and patent applications for all countries.



Explore the relationship between the percentage of female researchers and patent applications for the top 10 countries with high mean percentage female researchers.



ANOVA Test:

In different sectors, is there a difference in mean of scaled patent applications per capita?

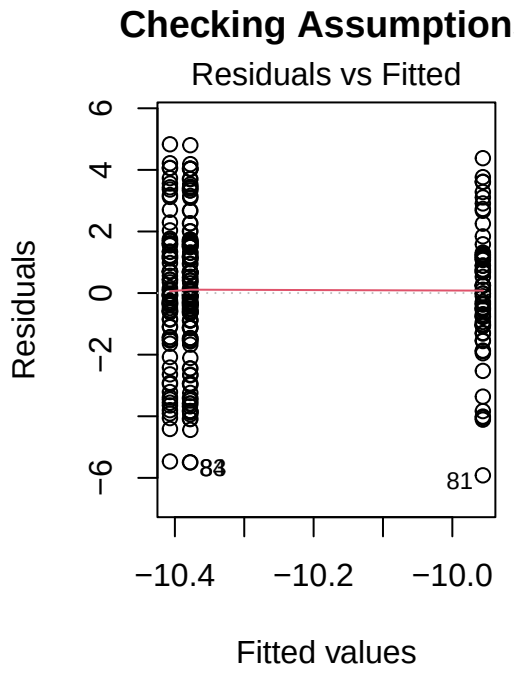
$\mu_1 = \mu_2 = \mu_3 = \mu_4$ are the means of scaled patent applications per capita for 4 sectors.

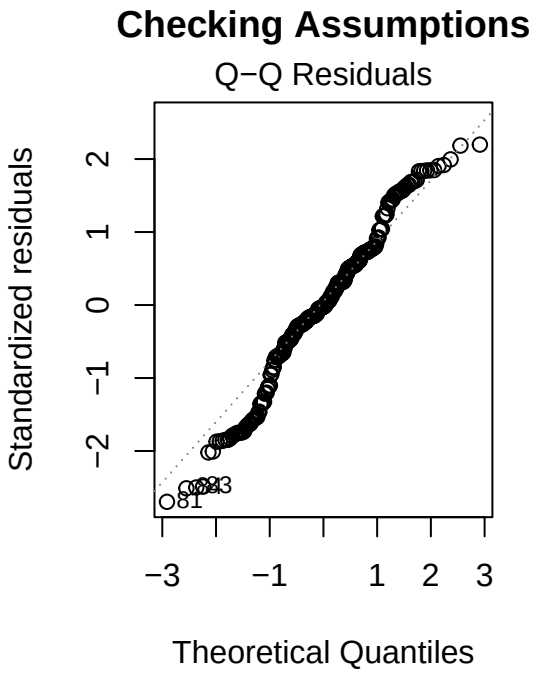
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$

$H_1 : \mu_i \neq \mu_j$ for at least one pair of i,j

Table 1: ANOVA Test Results(scaled patent applications per capita)

Term	Df	Sum.of.Squares	Mean.Square	F.Value	Pr..F.
Indicators	3	9.484	3.161	0.646	5.858789e-01
Residuals	275	1345.183	4.892	NA	NA





In different sectors, is there a difference in mean of percentage of female researchers?

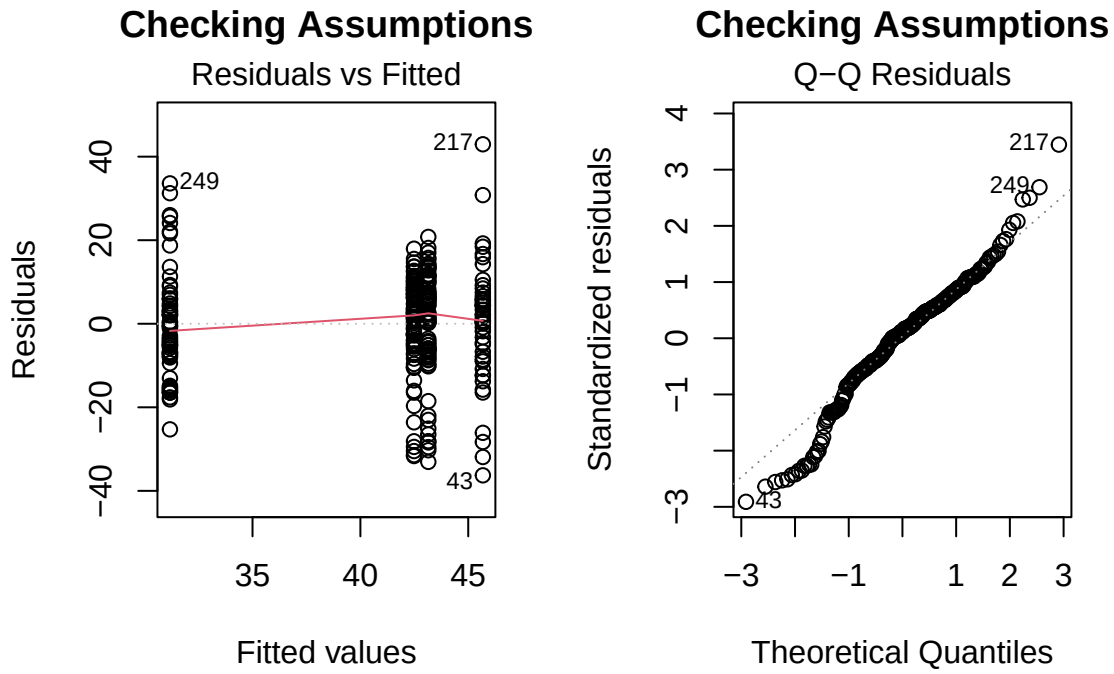
$\mu_1 = \mu_2 = \mu_3 = \mu_4$ are the means of percentage of female researchers for 4 sectors.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$

$H_1 : \mu_i \neq \mu_j$ for at least one pair of i,j

Table 2: ANOVA Test Results(mean percent of female researchers)

Term	Df	Sum.of.Squares	Mean.Square	F.Value	Pr..F.
Indicators	3	7923.336	2641.112	16.644	5.767621e-10
Residuals	275	43636.460	158.678	NA	NA



Correlation Test

Table 3: Correlation Test Results (Pearson's method)

Term	Estimate
Method	Pearson's product-moment correlation
Correlation Coefficient	0.133
t Statistic	2.232
P-value	0.026
95% Confidence Interval	0.016 to 0.247

Regression Model/ANCOVA

Simple linear Regression

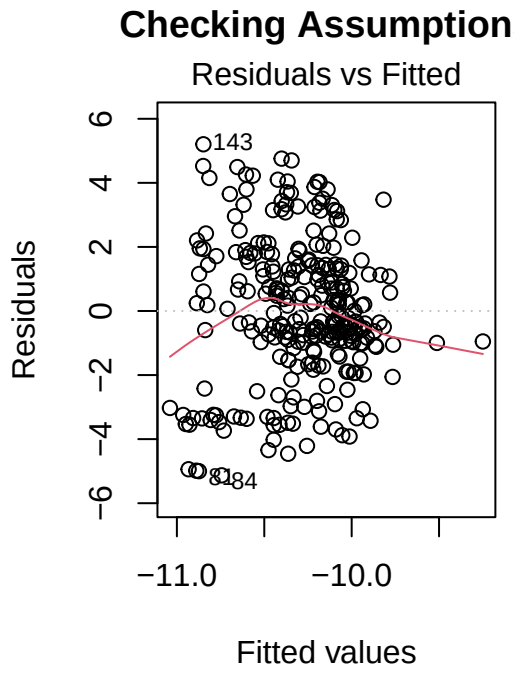
The graph shows that, in general, countries with more female researchers tend to have more patent applications. The horizontal axis represents the average percentage of female researchers across 4 sectors for each country from 2017 to 2023. The vertical axis represents the portion of patent applications. Each dot represents one country across 4 sectors.

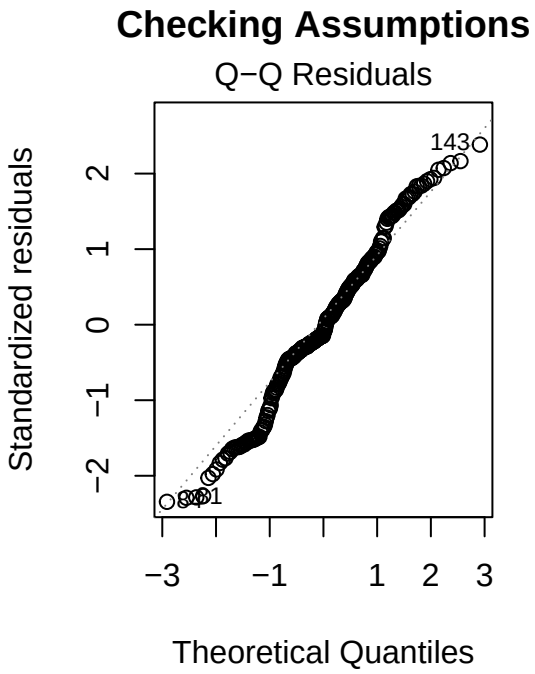
Table 4: Countries with More Than 75 Percent of Female Researchers

Country	Sectors	Median Scaled Patent Applications Per Capita	Mean Female Researchers (%)
Montenegro	Non-profit	-10.51	76.47
Serbia	Non-profit	-10.20	88.64

Table 5: Summary of Linear Regression Model

Term	Estimate	Std. Error	t value	p-value
Intercept	-11.1672	0.4146	-26.9351	0.0000
Mean Female Researchers	0.0216	0.0097	2.2322	0.0264

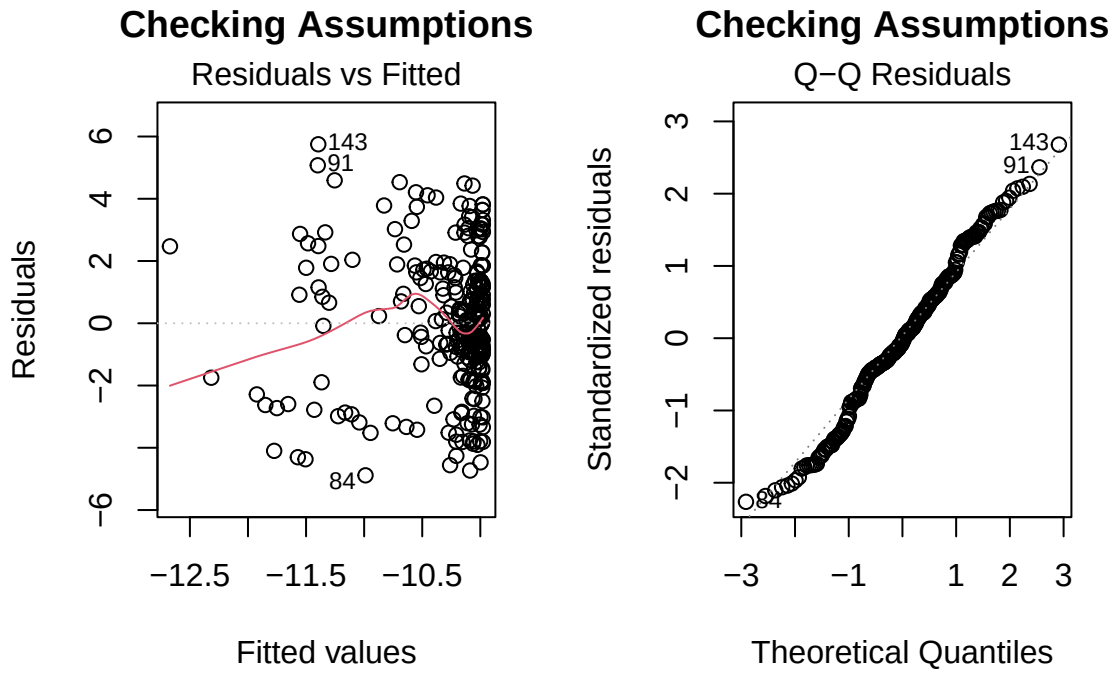




Polynomial Regression

Table 6: Summary of Polynomial Regression Model

Term	Estimate	Std. Error	t value	p-value
Intercept	-13.0578	0.7730	-16.8914	0.0000
Mean Female Researchers	0.1345	0.0403	3.3372	0.0010
Square of Mean Female Researchers	-0.0015	0.0005	-2.8827	0.0043



Influential points

Table 7: High Leverage and Outliers in the Polynomial Regression Model

Country	Sector	Median Patent Applications Per Capital	Mean Female Researchers (%)
Ethiopia	Business enterprise	-15.873897	13.55142
Ethiopia	Government	-15.873897	12.86531
Ethiopia	Higher education	-15.873897	10.82032
Germany	Business enterprise	-6.324169	14.68743
Luxembourg	Business enterprise	-5.644451	14.73494

Interaction term

Table 8: ANOVA Comparison of Models with Model Selection Based on AIC

model	AIC	Selected Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Linear Regression Model	1236.036	No	277	1342.203	NA	NA	NA	NA
Polynomial Regression Model	1229.760	No	276	1302.974	1	39.2299	9.2143	0.0026
Multiple Regression Model	1228.607	No	273	1269.989	3	32.9843	2.5824	0.0538

model	AIC	Selected Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Interaction Model	1206.803	Yes	270	1149.531	3	120.4581	9.4310	0.0000

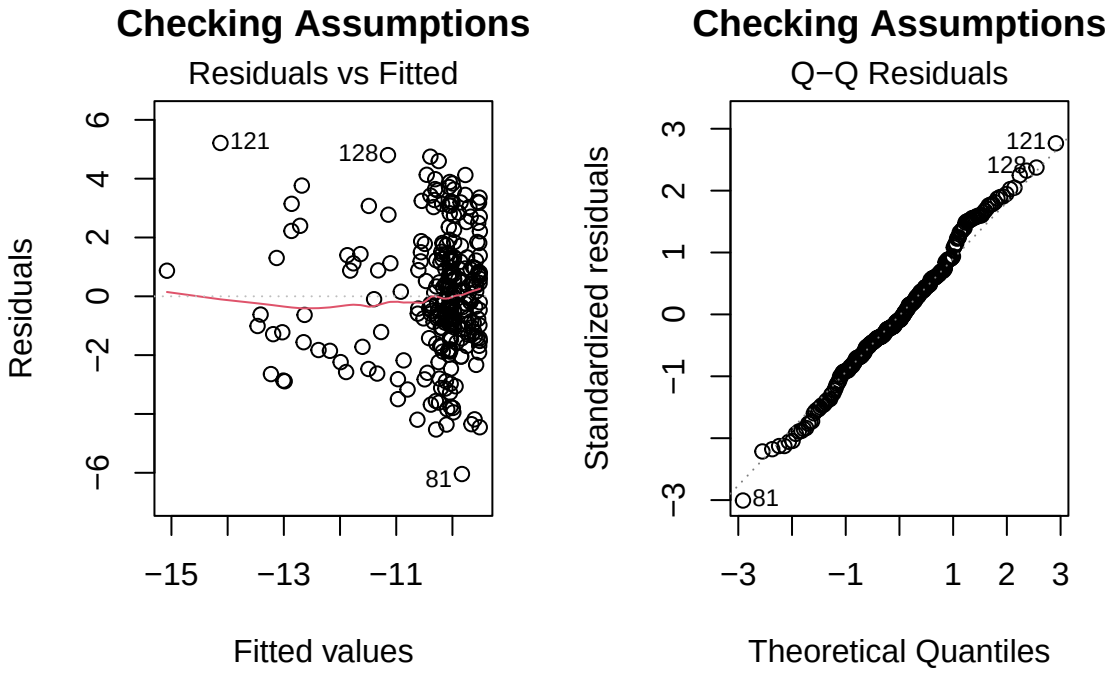


Table 9: Summary of the Interaction Model

	Estimate	Standard Error	t Value	p Value
(Intercept)	-10.9941	0.8929	-12.3123	0.0000
Mean Female Researchers	0.1172	0.0429	2.7331	0.0067
Squared Mean Female Researchers	-0.0023	0.0005	-4.3386	0.0000
Indicators (Government)	-4.5263	1.0708	-4.2270	0.0000
Indicators (Higher education)	-4.3718	1.1181	-3.9101	0.0001
Indicators (Non-profit)	-6.4249	1.2713	-5.0538	0.0000
Mean Female Researchers:Indicators (Government)	0.1097	0.0274	3.9959	0.0001
Mean Female Researchers:Indicators (Higher education)	0.1053	0.0286	3.6766	0.0003
Mean Female Researchers:Indicators (Non-profit)	0.1535	0.0309	4.9627	0.0000

Table 10: High Leverage and Outliers in the Interaction Model

Country	Sector	Scaled Patent Applications Per Capita(Median)	Percentage of Female Researchers
Iran (Islamic Republic of)	Non-profit	-8.9142	13.7825

Summary

The ANOVA test revealed that there is significant variation in the percentage of female researchers across different sectors. Specifically, the business enterprise sector has a lower percentage of female researchers, which correlates with a decreasing trend in patent applications. On the other hand, in the government sector, there is no obvious relationship between the percentage of female researchers and patent applications. The higher education and non-profit sectors showed a positive relationship between female researchers and patent applications, suggesting that higher female representation could contribute to increased innovation in these sectors.

The correlation test comparing the mean percentage of female researchers and the median of scaled patent applications per capital from 2017 to 2023 indicated a slight positive linear relationship. However, this relationship is not uniform across sectors, pointing to the importance of considering sector-specific factors when analyzing the connection between female researchers and patent innovation. Countries with a higher percentage of female researchers exhibited a stronger positive relationship between female researchers and patent applications, further reinforcing the idea that higher female representation may lead to more innovation.

The regression analysis showed that the relationship between female researchers and patent applications is not straightforward. Simply increasing the percentage of female researchers does not always result in more patents. Without considering specific sectors, only about 1.5% of the variability in patent applications is explained by the percentage of female researchers. However, when sectors are included, the explanatory power increases to about 5%, emphasizing the significance of sector-specific factors. The impact is nonlinear, with the effect varying depending on the sector. In this analysis, sectors matter—the relationship between female researchers and patent applications changes depending on the sector. For example, government, higher education, and non-profit sectors show different patterns compared to the business enterprise sector.

Finally, the interaction term in the regression model indicated that the impact of female representation on patent applications depends on the sector. The interaction between mean female researchers and sectors reveals that the influence of female researchers on patent applications is not uniform across all sectors. For instance, the impact is stronger in the higher education and non-profit sectors, while it is weaker in the business enterprise and government sectors. This highlights the importance of sector-specific strategies to encourage female representation and innovation.

Conclusion

The result shows encouraging female researchers can drive innovation, but the sector is an important factor. Further research is necessary to understand why certain sectors benefit more from female participation. Specialized strategies needed to support females in research to foster innovation and patent success.