

# Job Postings: Experience vs Salary by Data Language

## Introduction

This short report examines a sample of job postings to explore the relationship between required minimum years of experience and reported median salary, and to compare salary distributions by declared data languages (R, Python, Both, Neither).

## Scatter plot: Minimum Years Experience vs. Median Salary



## Brief interpretation

The scatter plot above displays the relationship between minimum years of experience and the reported median salary. A fitted linear trend suggests a modest positive relationship (higher experience tends to link with higher salary), but there is substantial variance at most experience levels. Inline summary: median salary = \$68,000, observations plotted = 400.

## Box plot: Median Salary by Data Languages



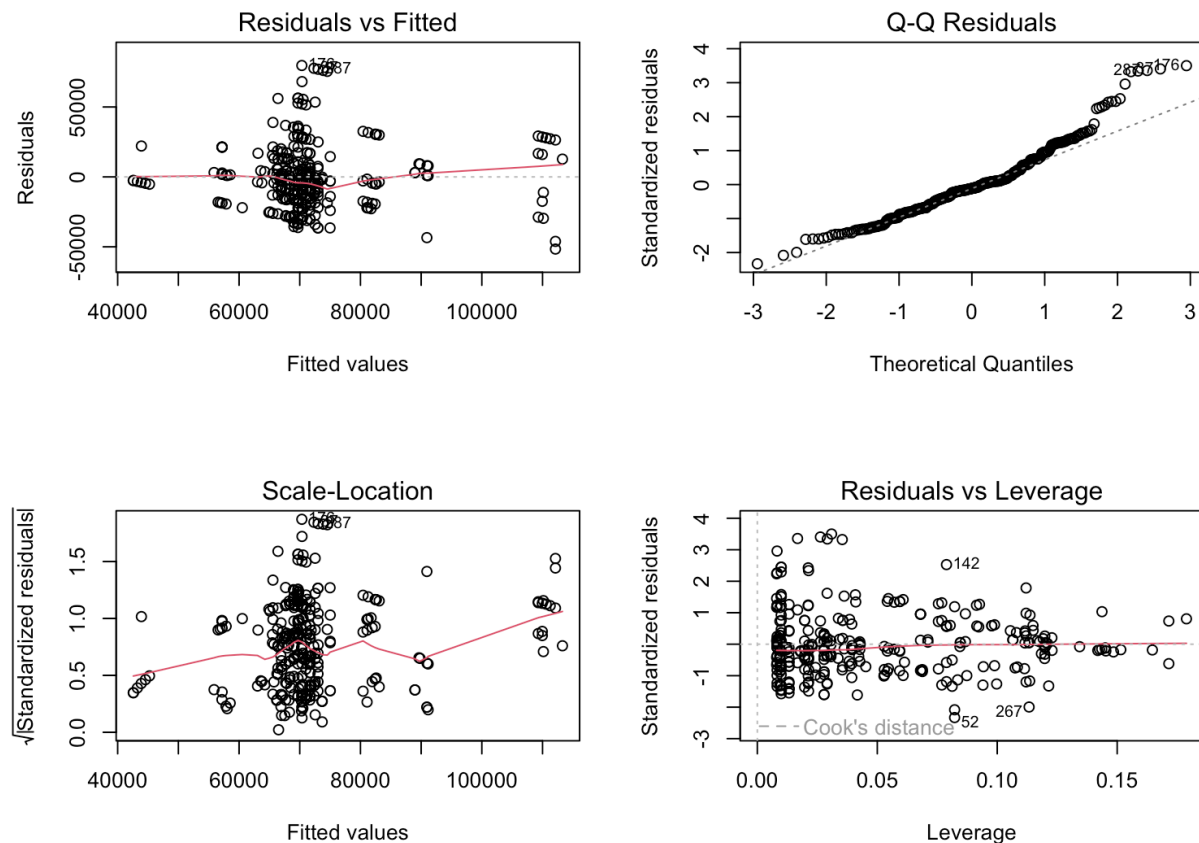
### Brief interpretation

The box plot compares salary distributions for jobs that mention R, Python, both, or neither. In this sample, postings that list both R and Python often display higher median salaries and higher upper quartiles, while postings listing neither show a wider spread including lower medians. However, distributions overlap and sample sizes per group may vary.

### Conclusion

- There is a modest positive association between minimum years of experience and median salary in this sample, but variation is large at each experience level.
- Jobs that list both R and Python tend to have higher median salaries in this dataset, though overlap and sample size differences mean the result is descriptive only.
- For stronger inference, expand the dataset and control for role, industry, location, and company size.

## Diagnostics plots



## Interpretation of regression

The linear model predicts median salary from minimum years, declared data language, and a simplified location grouping. Key outputs above include estimated coefficients (with standard errors and p-values), R-squared, and VIFs to check multicollinearity. Residual and influence plots (shown) help assess linear model assumptions: linearity, homoscedasticity, and influential observations.

## Reflection on the process

While using Cursor's Agent mode, I was struck by how quickly AI tools are progressing. However, it's clear that I still need to know what I'm doing as the user. The tool itself was very easy to use. Typing R code has always been a hurdle for me. With Cursor, I don't have to write the code myself; instead, my responsibility shifts to checking whether the output is correct. This shows that asking the right questions and using the tool the right way is what matters most.