Question 3: Did sex bias exist in granting salary increases between 1990 and 1995?

Answer 3:

To address this question, our team developed an interactive R Shiny application complemented by this detailed analysis.

Methodology:
We first consolidated employee IDs and filtered for those consistently present between 1990 and 1995. To avoid double-counting salary increases, we calculated the average salary increment per individual over this five-year period. This refined dataset was subsequently used to produce visualizations and develop linear regression models.

Findings from Linear Regression Models:

Gender as Sole Predictor:
A linear regression model with gender as the only predictor yielded an intercept of 212 and a coefficient (slope) for males (sexM) of 48.518. The p-value associated with this coefficient was extremely low (7.45e-12), indicating strong statistical evidence of gender-based salary increment disparities. However, it's important to note that the absolute difference in increments between men and women (approximately $48.52) is modest, especially when averaged over five years.

Despite the statistical significance of gender, this model had an adjusted R^2 value of only 0.03. This indicates that while gender is statistically significant, it explains very little (only 3%) of the variability in salary increments.

```
Call:
lm(formula = avg_increment ~ sex, data = salary_data)

Residuals:
    Min      1Q  Median      3Q     Max
-260.96  -79.26  -13.33   57.63 1157.67

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  212.009      6.090  34.811  < 2e-16 ***
sexM          48.518      7.029   6.903 7.45e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 118.7 on 1523 degrees of freedom
Multiple R-squared:  0.03034,   Adjusted R-squared:  0.0297
F-statistic: 47.65 on 1 and 1523 DF,  p-value: 7.451e-12
```

Expanded Predictors Model:
Our R Shiny App enabled further exploration with additional predictors: gender, year of degree obtained, rank, administrative role status, highest degree held, and the field of work. Incorporating all six predictors without interaction terms improved the adjusted R^2 value to 0.365. Although this model significantly improved explanatory power compared to gender alone, it still failed to explain approximately two-thirds of the variability in average salary increments.

Interaction Terms Analysis:
Introducing interaction terms (up to 6-way interactions) did not significantly enhance model performance. The best interaction model tested, avg_increment ~ (yrdeg + rank + admin + field)^6, yielded an adjusted R^2 of approximately 0.36, marginally lower than the simpler model (adjusted R^2 of 0.365).

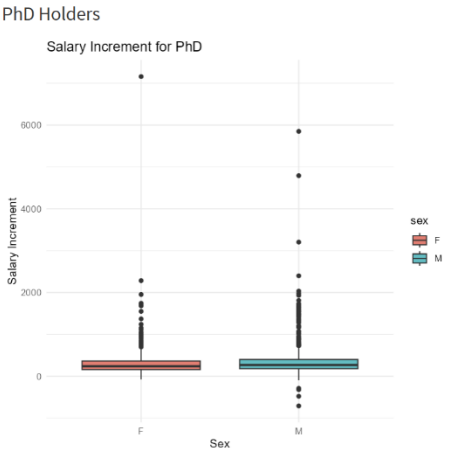
	Model
53	avg_increment ~ (yrdeg + rank + admin + field)^6
62	avg_increment ~ (yrdeg + rank + admin + deg + field)^6
58	avg_increment ~ (sex + yrdeg + rank + admin + field)^6
63	avg_increment ~ (sex + yrdeg + rank + admin + deg + field)^6
49	avg_increment ~ (sex + rank + admin + field)^6

	Predictors	R2	Adj_R2
53	yrdeg, rank, admin, field	0.3719385	0.3593268
62	yrdeg, rank, admin, deg, field	0.3886979	0.3592680
58	sex, yrdeg, rank, admin, field	0.3837907	0.3585362
63	sex, yrdeg, rank, admin, deg, field	0.4091561	0.3568242
49	sex, rank, admin, field	0.3635068	0.3507258

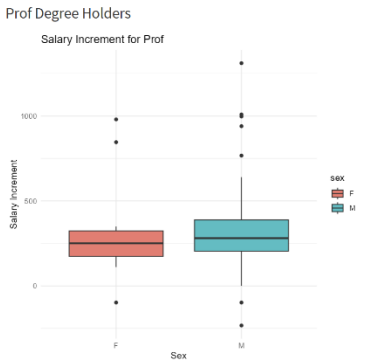
Now, with that being said, when we look at various particular visualizations of the data, we find some interesting patterns. These are discussed in points below.

Analysis by Degree Type:

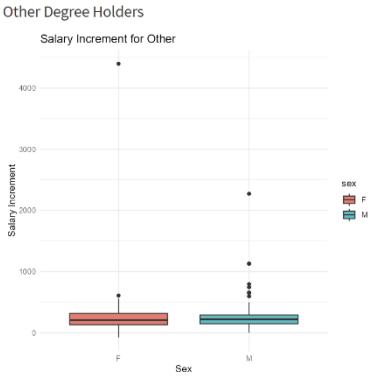
- **PhD Degree Holders:**
    - No significant gender gap was observed in average salary increments between men and women. However, due to the higher number of men in the sample, male outliers were more frequent, though their increment values were only marginally higher.



- **Professional Degree Holders:**
    - Men exhibited a higher average salary increment compared to women, particularly noticeable at the 75th percentile, indicating that top earners among men had significantly larger increases.

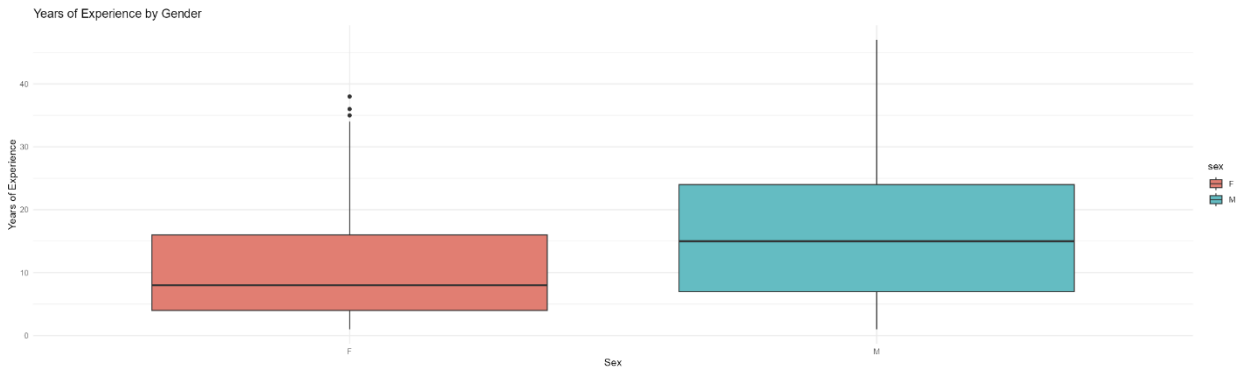


- **Other Degree Holders:**
    - Average salary increments were similar for both genders, but the 75th percentile favored women slightly. However, the greater number of male outliers could potentially skew this comparison.



Experience and Salary Increment Analysis:

Years of Experience by Gender

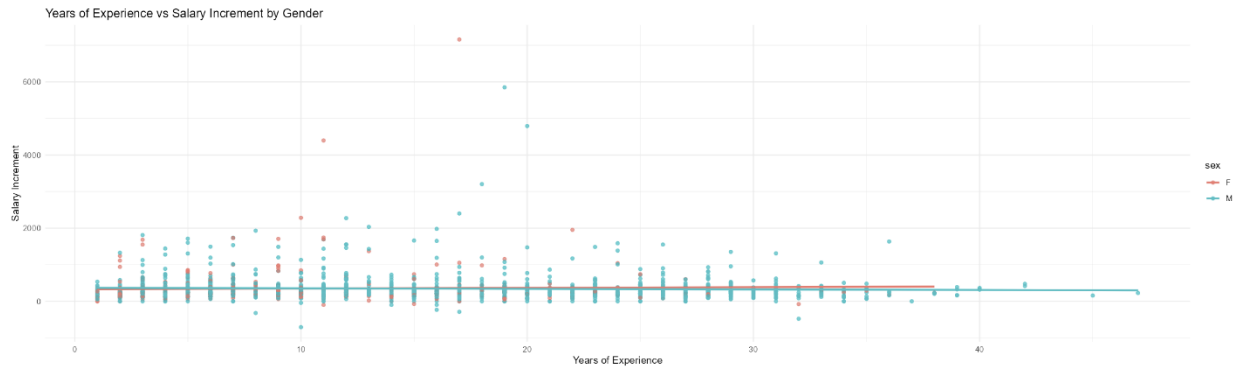


- During the study period (1990-1995), average years of experience for female faculty were approximately 8 years, whereas for male faculty it was roughly 14.5 years. This difference of over six years in experience likely influenced salary increment disparities.

- A regression model using average increment as a response variable and sex and year of degree (a proxy for experience) as predictors with an interaction term (avg_increment ~ (sex + yrdeg)^2) revealed sex as statistically insignificant (p = 0.553), while year of degree remained highly significant (p = 2e-07). The interaction term was also statistically insignificant (p = 0.287), and the model itself had limited explanatory power (R^2 = 0.082).

- A visual analysis of salary increments against years of experience indicated no clear linear relationship. However, it highlighted a notable pattern of higher increments awarded disproportionately to men, irrespective of experience.

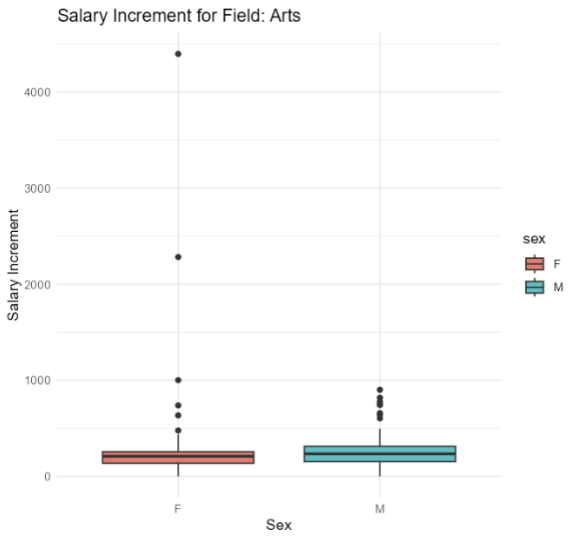
Years of Experience vs Salary Increment



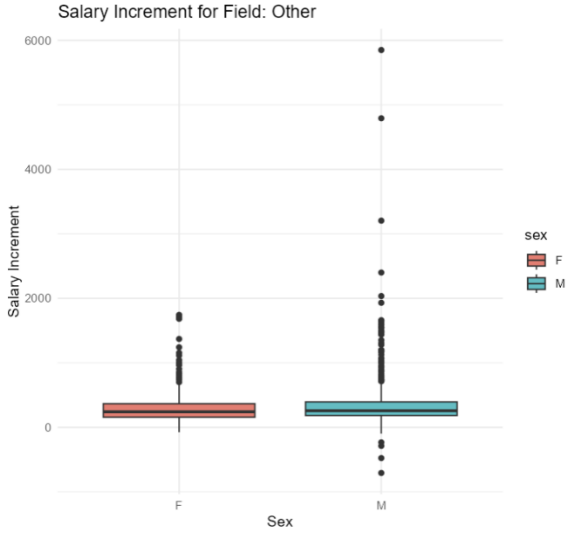
Salary Increment Analysis by Field:

- **"Other" Field:**
  - Average salary increments for men and women were similar, though male outliers were more frequent.
- **Arts Field:**
  - Increment means for both genders were comparable, with outliers being almost equally distributed.
- **Professional Field:**
  - Although average increments for men and women were nearly identical, the 75th percentile favored women. Still, the presence of more male outliers indicated substantial individual variability.

Field: Arts



Field: Other



Field: Prof