Graded Team Assignment - Gender Discrimination Lawsuit

Group 11

Agenda

1. Conclusion
2. Key Claims
3. Key Claims - Salary
4. Key Claims - Rank

Conclusion

Conclusion

1. **Female doctors' have 2 key claims: Salary Disparity and Glass Ceiling**

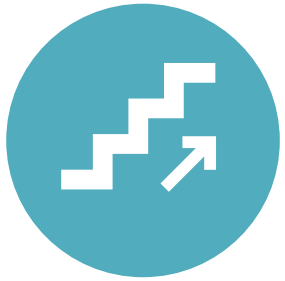
2. **Based on both descriptive statistics and linear regression, females have lower salary performance compared to males.**

3. **From both modelling and visualization standpoints, females' chance for climbing up corporate ladder is evidently low.**

4. **Since both key claims are justifiable, we conclude that gender discrimination exists.**

Key Claims

Female Doctors' Key Claims: Salary Disparity and Glass Ceiling

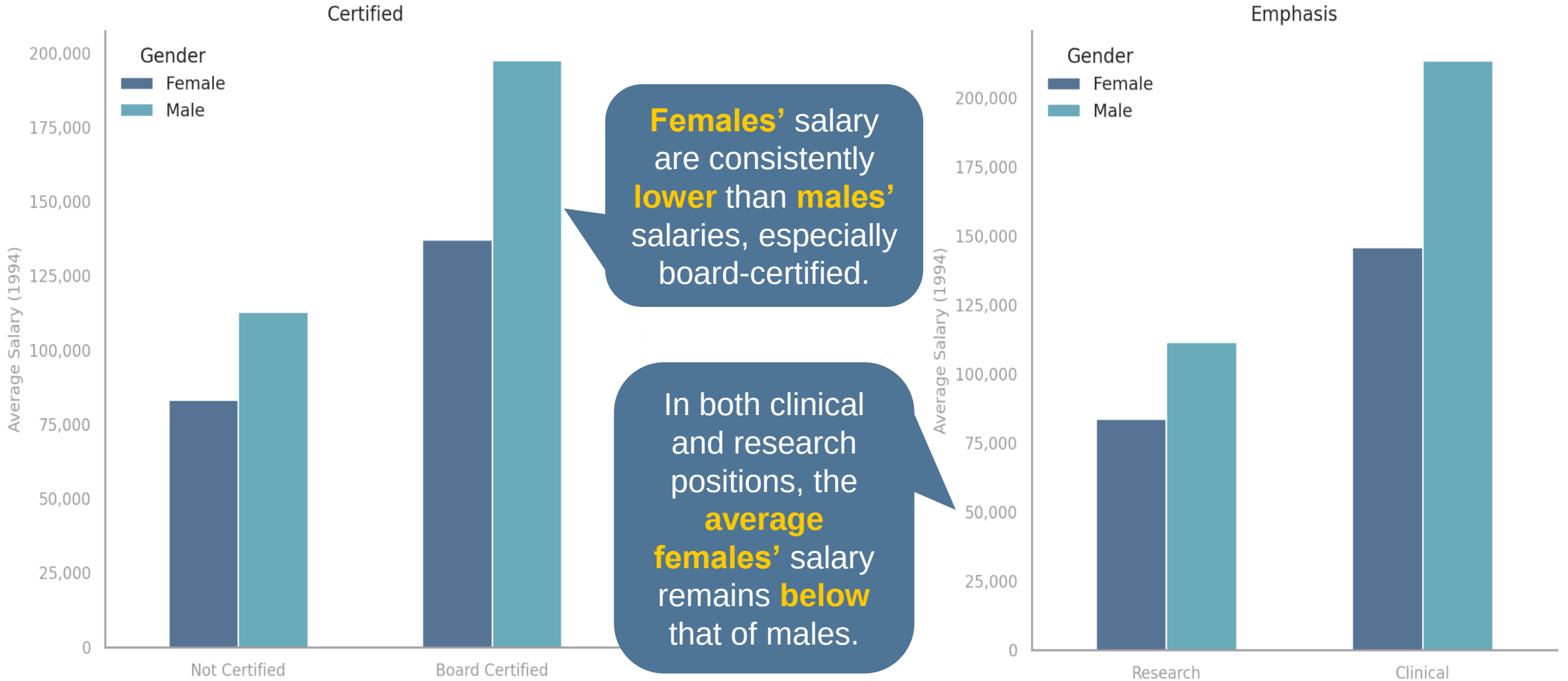


Claims	Salary Disparity	Glass Ceiling
Arguments	Women earn less money than men on average	Female faculty were less likely to be full professors
Target variables	Salary 1994, Salary 1995	Rank

Shih Hsiao Ju

Key Claims - Salary

Females' Salary Performance is Weaker Across All Attributes



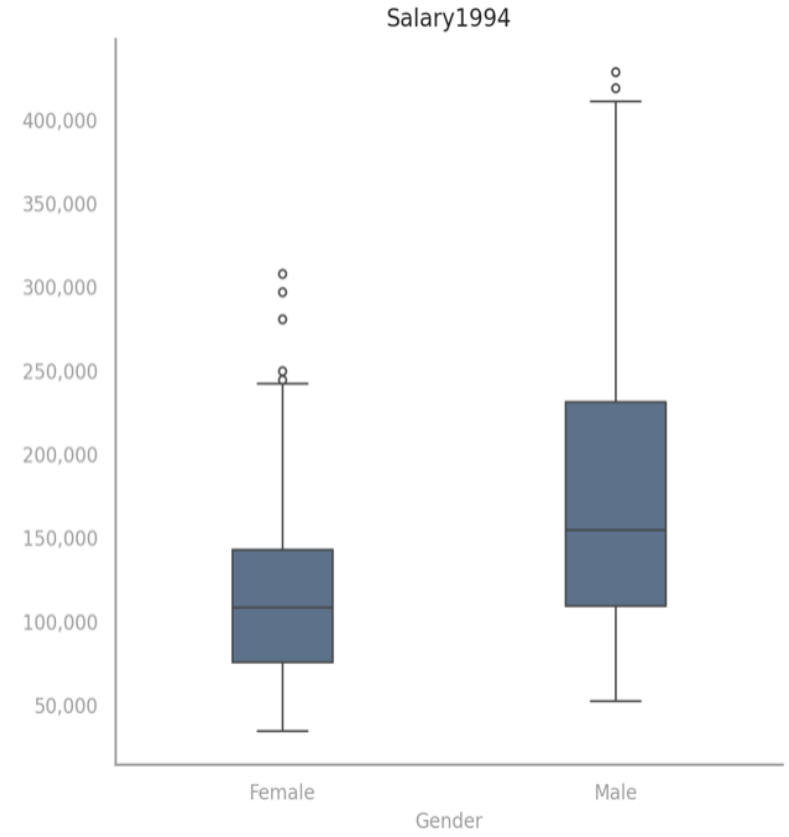
Wu Sin Pei

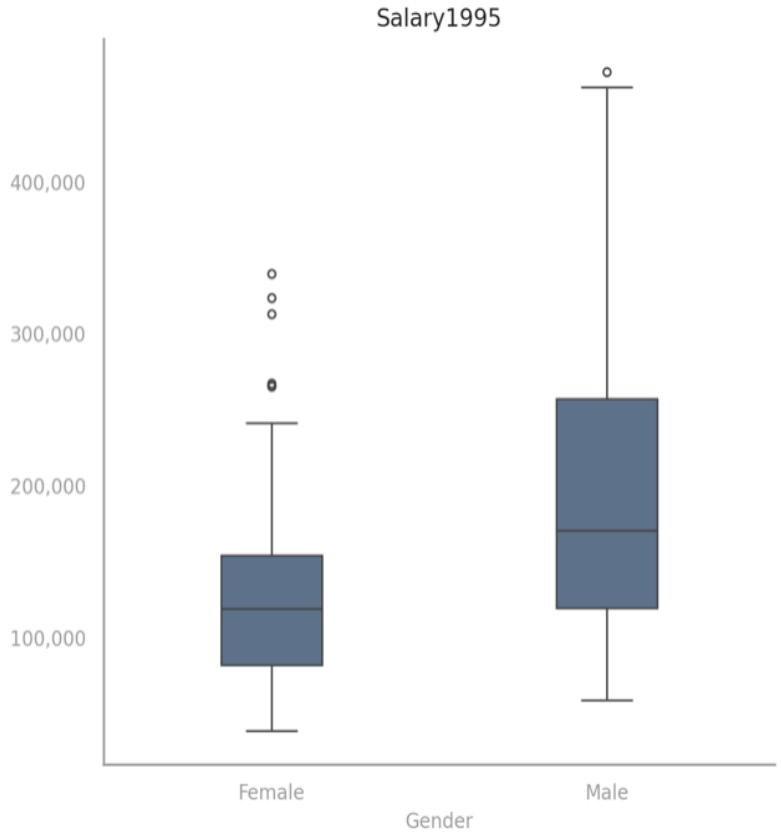
Females' Salary Performance is Weaker Across All Attributes

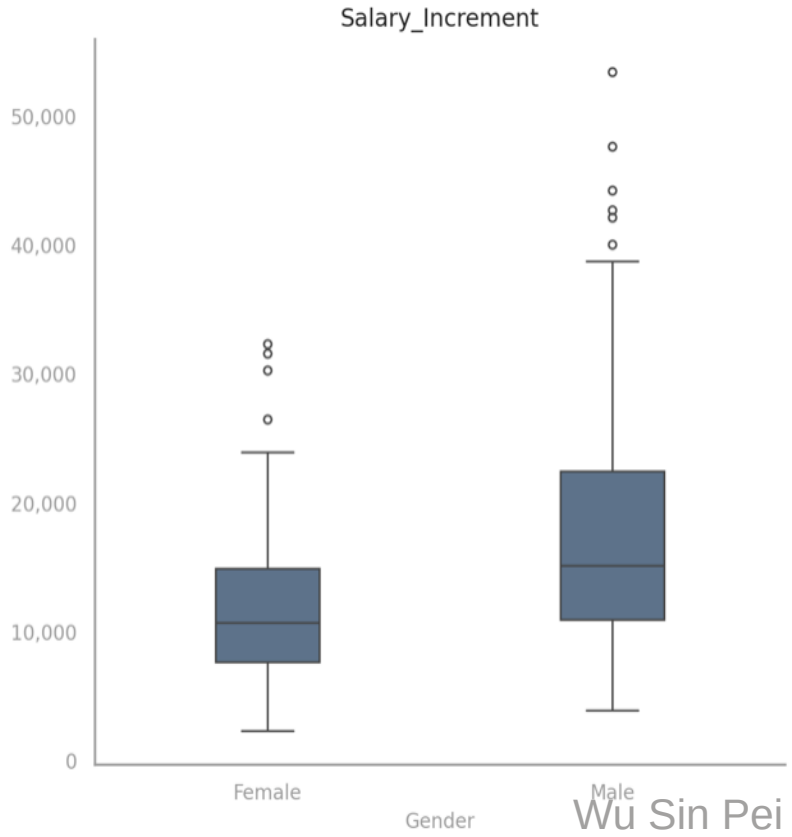
Females had a narrower IQR, and the outliers are lower than the males, indicating a **significant pay gap** and less salary variability.

Females' **average salary remained lower** than males' in 1995, with persistent low-salary outliers for females.

Females saw **smaller salary increases** than males, with lower mean increments, less variability and fewer outliers.

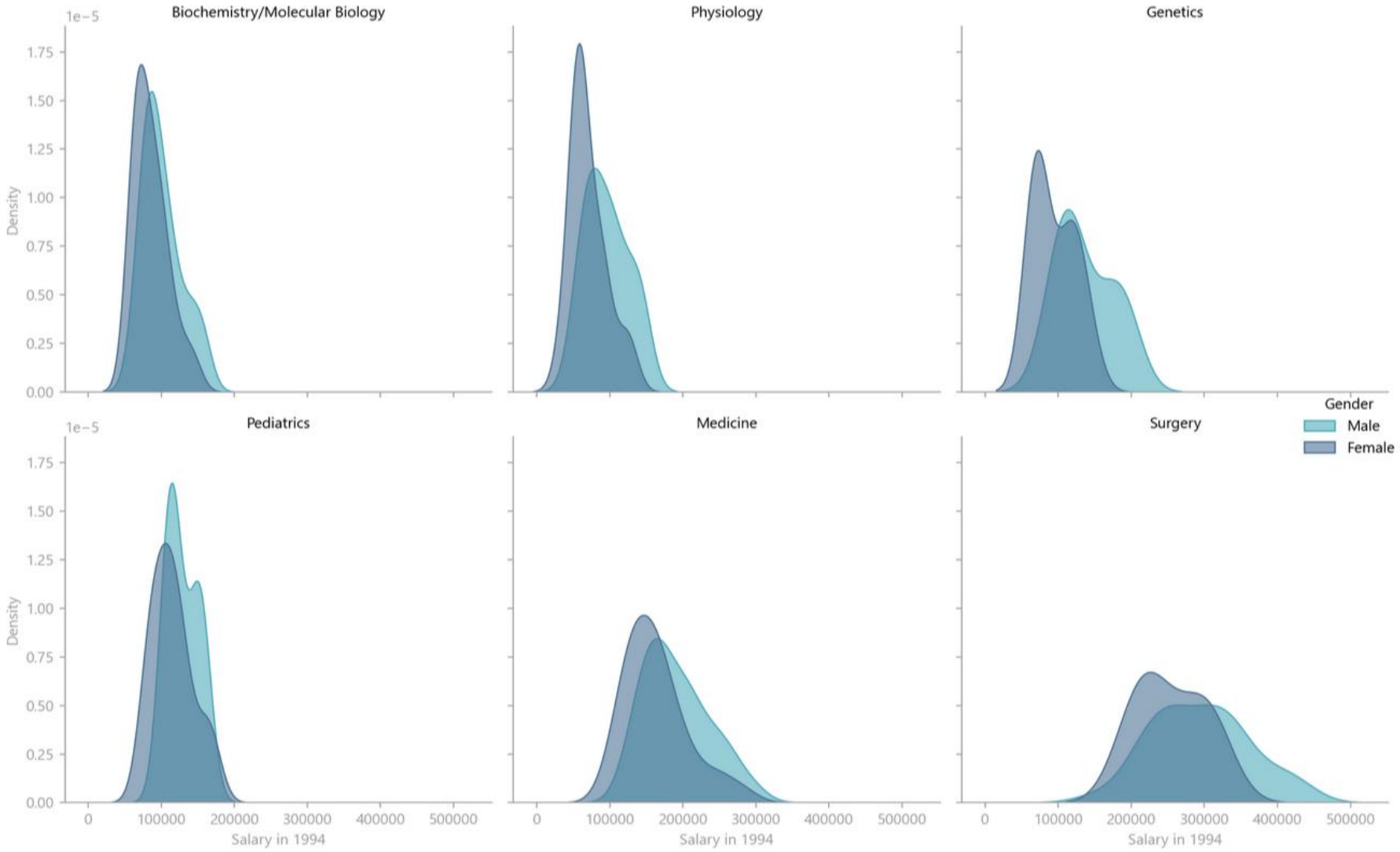






Wu Sin Pei

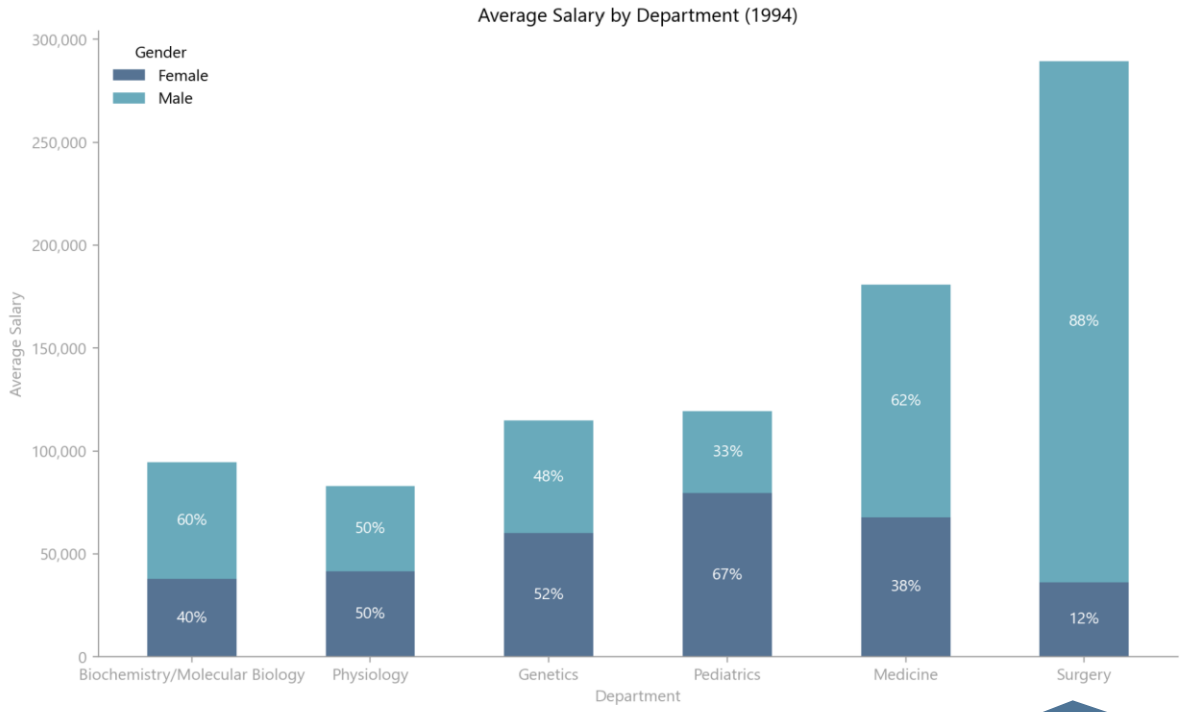
Females' Salary is Lower Than Males' Across All Departments

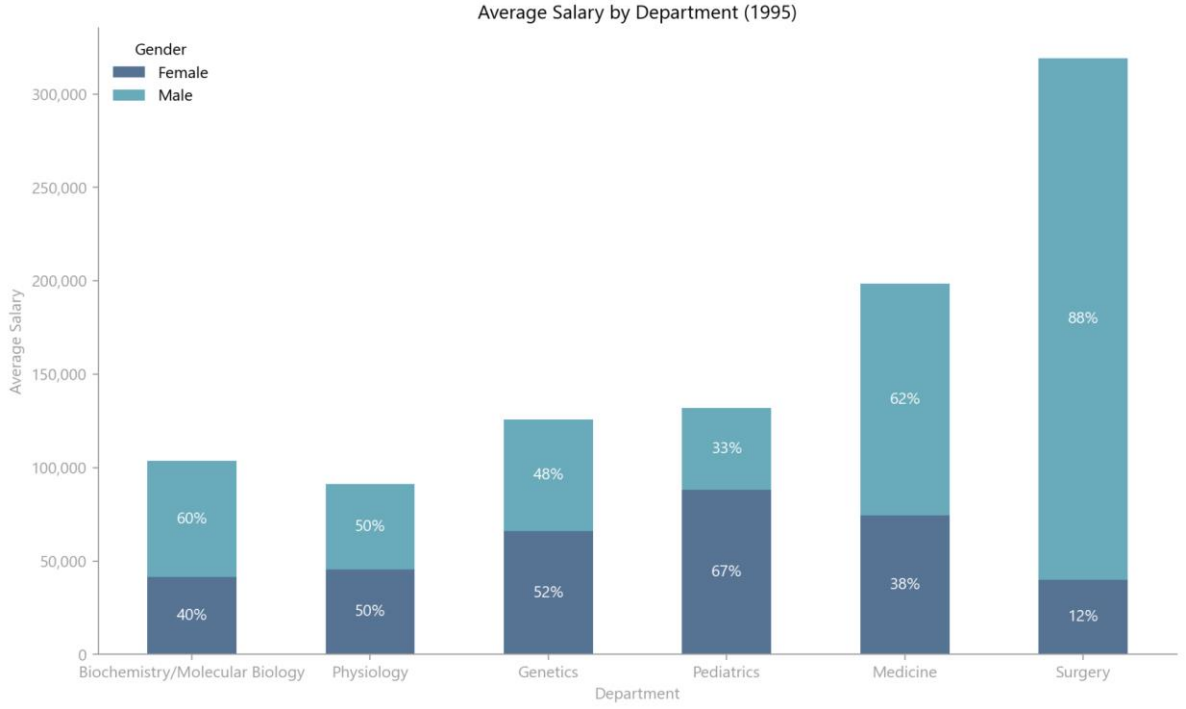


Males' salaries are much higher than **Females'** in each department.

Cao Yimeng

Females are Unable to Join the Highest-paid Department





Surgery department has the **highest** average salary and the **lowest** female proportion in 1994 and 1995.

Zhang Jing

Linear Regression: Males have higher salaries than Females

	Constant	Gender	Clin	Cert	Log_Prate	Log_Exper	Physiology	Genetics	Pediatrics	Medicine	Surgery	Associate	Full Professor
Log_Sal94	10.93^{***}	**0.02**	0.18^{***}	0.18^{***}	-0.09	0.18^{***}	-0.17^{***}	0.16^{***}	0.17^{***}	0.51^{***}	0.87^{***}	0.11^{***}	0.21^{***}

Based on our explorations in the later rank section, we found a strong correlation between **Rank** and **Gender**

To minimize the effect of multicollinearity, we chose to **remove Rank** from features

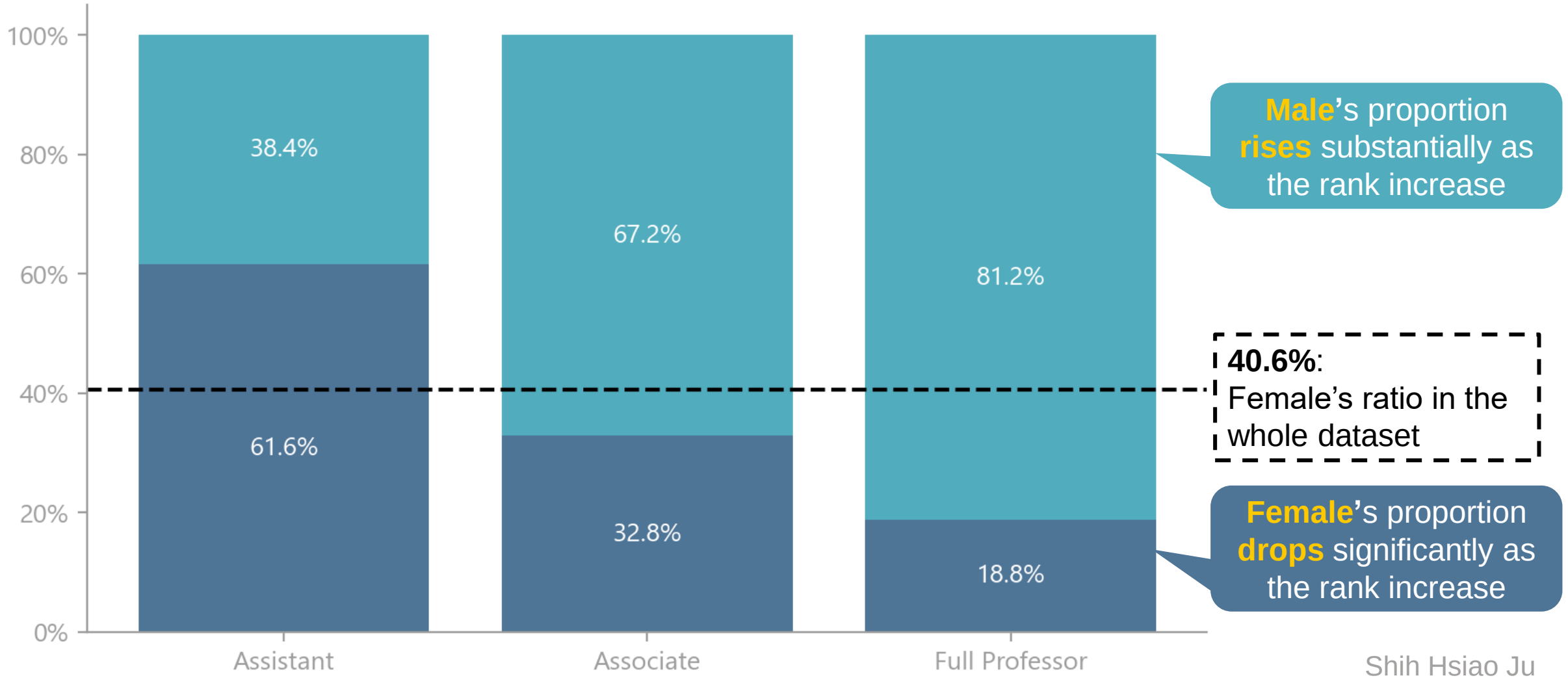
	Constant	Gender	Clin	Cert	log_Prate	Log_Exper	Physiology	Genetics	Pediatrics	Medicine	Surgery
Log_Sal94	10.85^{***}	$\mathbf{0.05}^{**}$	0.18^{***}	0.17^{***}	-0.11	0.27^{***}	-0.15^{***}	0.16^{***}	0.16^{**}	0.50^{***}	0.83^{***}

** represents the level of significance, the more *, the more significant .*

Cao Yimeng

Key Claims - Rank

The Percentage of Females as Full Professor is Apparently Low



Logistic Regression: Males have a higher likelihood of promotion

$$Logit(p) = \ln(\frac{p}{1-p}) = \beta_0 + \beta_1 \cdot Gender + \beta_2 \cdot Exper + \beta_3 \cdot Prate + \beta_4 \cdot Clin + \beta_5 \cdot Cert$$

- p = P(Promoted=1): This represents the probability that a person is promoted (Rank = Professor or Associate)
- $1-p$ = P(Promoted=0): This represents the probability that a person is not promoted (Rank = Assistant).

	Constant	Gender	Exper	Prate	Clin	Cert
Logit(p)	-4.1751^{***}	1.1556^{**}	0.4957^{***}	0.035	-0.0571	-0.9372^{*}

** represents the level of significance, the more *, the more significant .*

Logistic Regression: Males have a higher likelihood of promotion

Gender

Gender has a **significant effect** on promotion. The P-value is 0.003, and the coefficient is 1.156, indicating that males are more likely to be promoted.

Experience (Exper)

Work experience also has a **significant impact** on promotion, with a P-value of 0.000 and a coefficient of 0.495.

Certification (Cert)

Certification **has some influence**, but it is slightly below the significance level. The P-value is 0.041, with a coefficient of -0.937.

Appendix

CART's Performance is not impressive

Professor ~ PublicationRate + Experience+Salary1994+Salary1995+SalaryIncre+Dept+Gender+Emphasis+Certified

Nodes	Depth	Score	STDERR
3	1	0.766	0.090
5	2	0.751	0.088
9	4	0.759	0.093
11	4	0.747	0.084
21	8	0.747	0.084
31	11	0.747	0.085
33	11	0.751	0.089
35	11	0.748	0.086
41	11	0.732	0.098
43	11	0.732	0.097
43	11	0.732	0.097
43	11	0.732	0.097
51	13	0.732	0.097
53	13	0.728	0.098
57	13	0.713	0.090
61	13	0.724	0.078
61	13	0.724	0.078
69	13	0.724	0.078
73	13	0.724	0.078
81	13	0.724	0.078
85	13	0.724	0.078
91	13	0.724	0.078
95	13	0.724	0.078
103	13	0.724	0.078

Best tree (based on 1 stderr rule) only score 76.6%, **slightly higher than the majority rule (67%)**

As the tree only splits once, the only feature affecting the model is Experience (x[1]). However, since the model performs poorly, we will suggest the result is **not persuasive**.