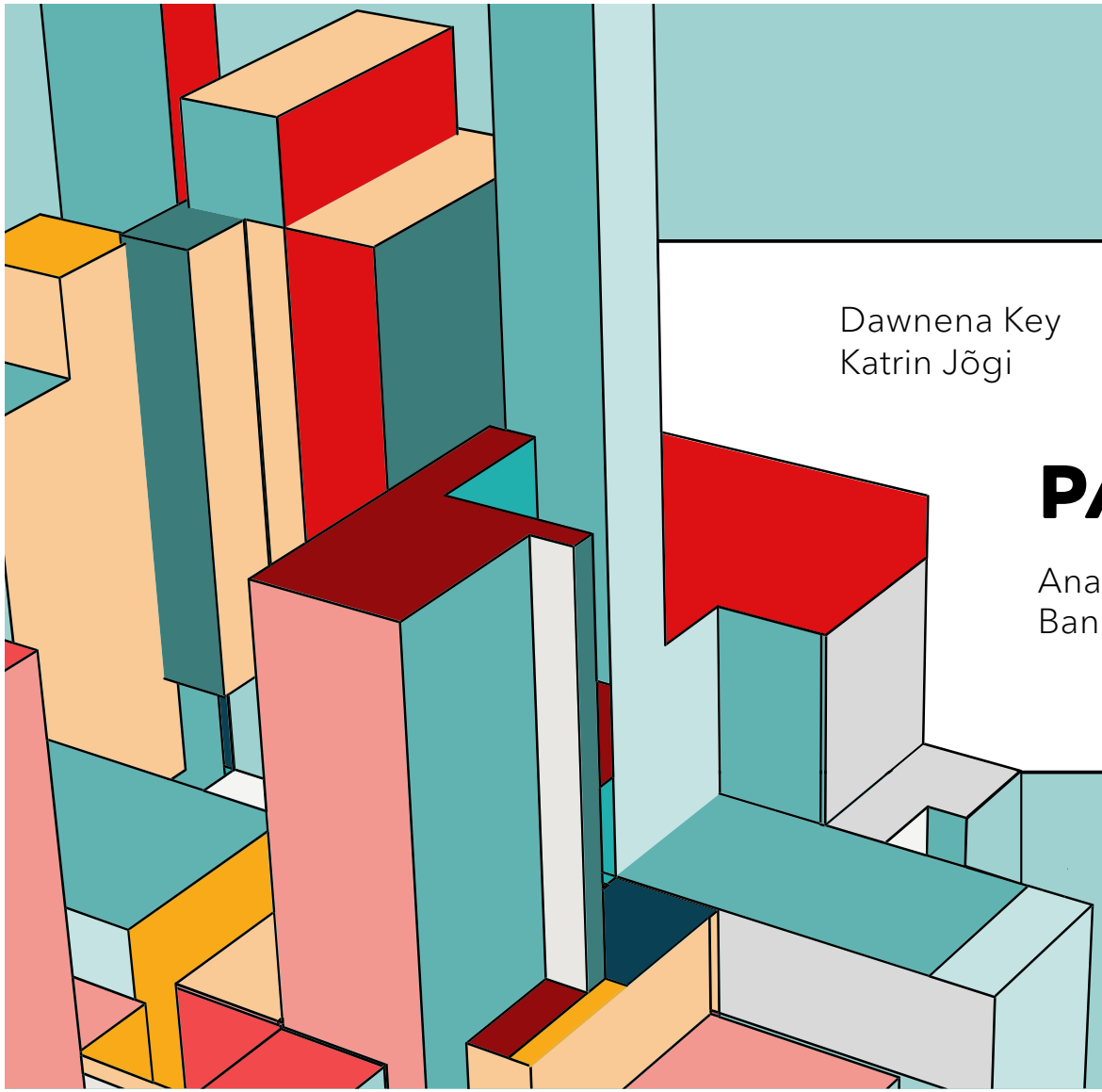


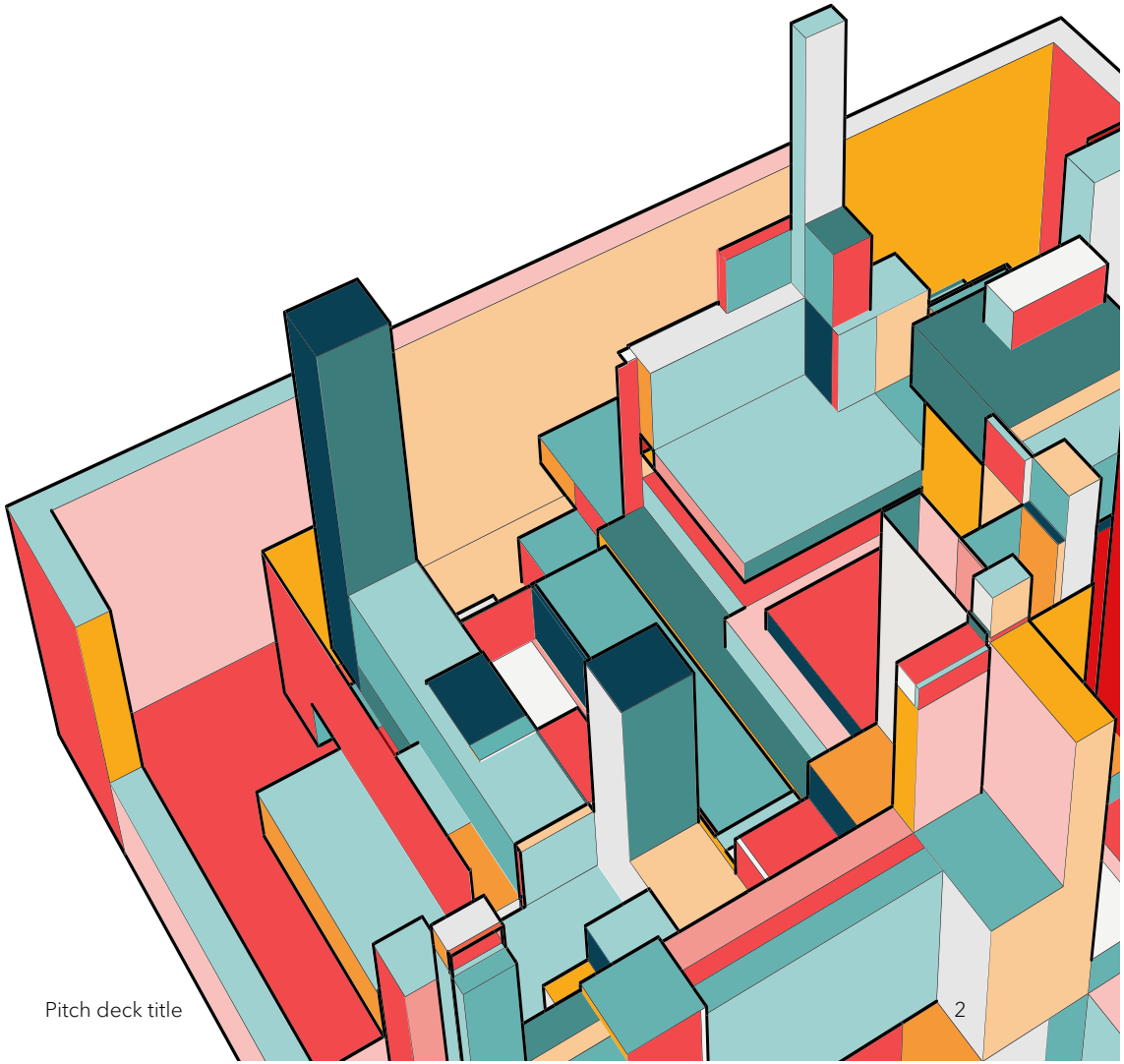
Dawnena Key
Katrin Jõgi

PAY DISPARITY

Analysis of Gender Pay Disparity at a Bank

AGENDA

Background
Data Overview
Statistical Methods
Limitations & Conclusions



BACKGROUND

ISSUE

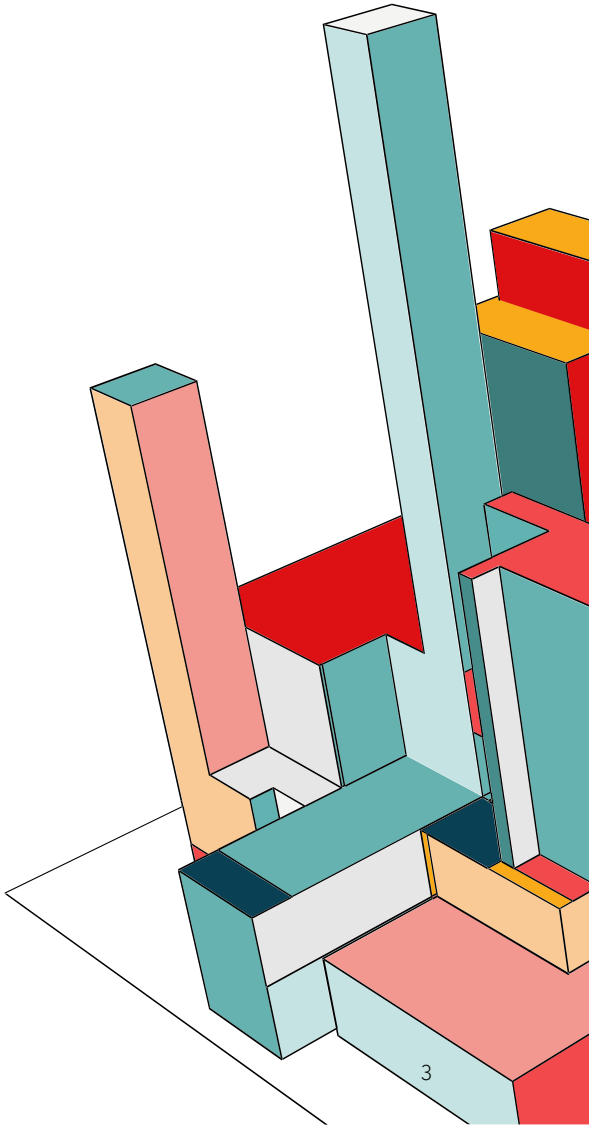
Potential gender-based discrimination at bank

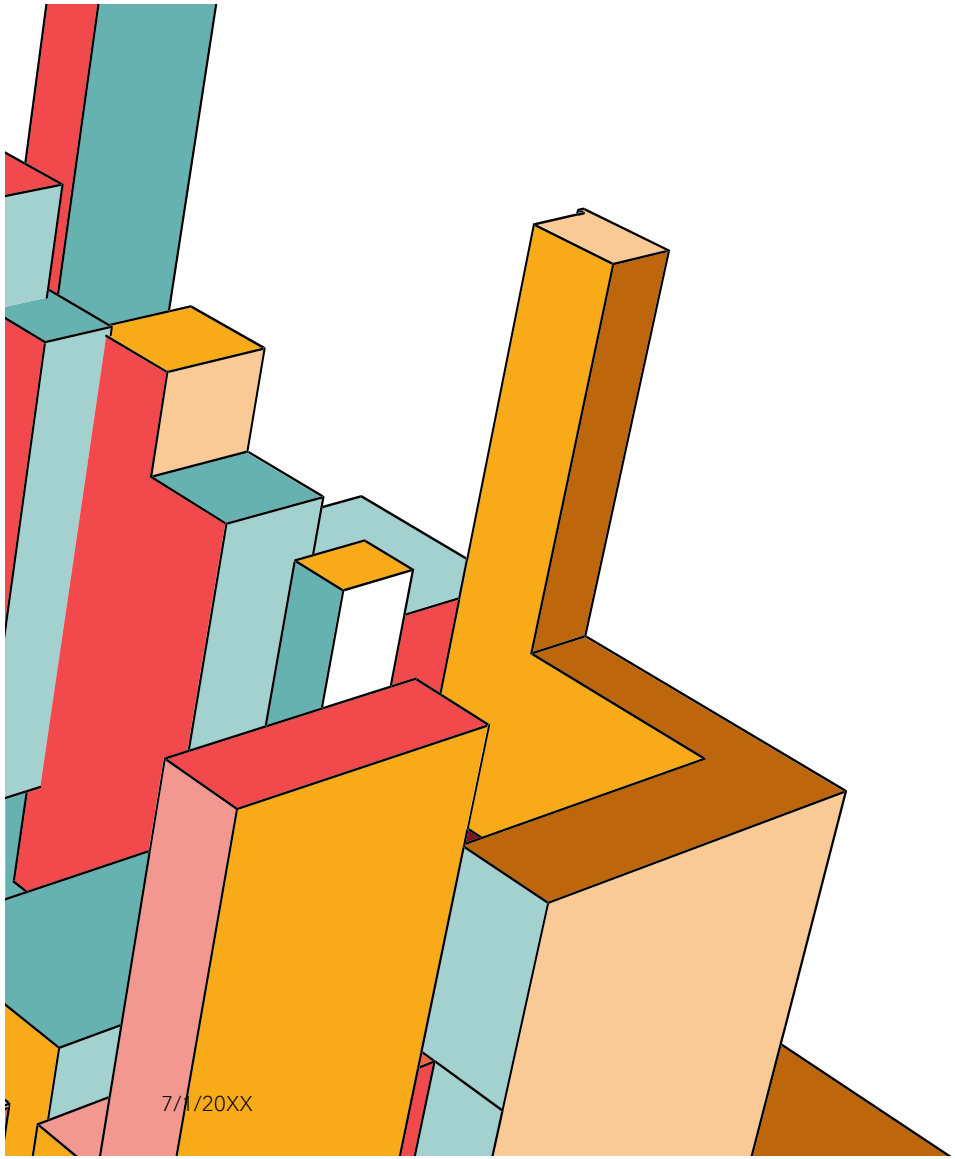
DATA

Data on skilled, entry-level clerical employees of a bank

GOAL

Use statistical methods to test for gender-based pay disparity at a bank





DATA OVERVIEW

VARIABLES

Response variable: Starting salary

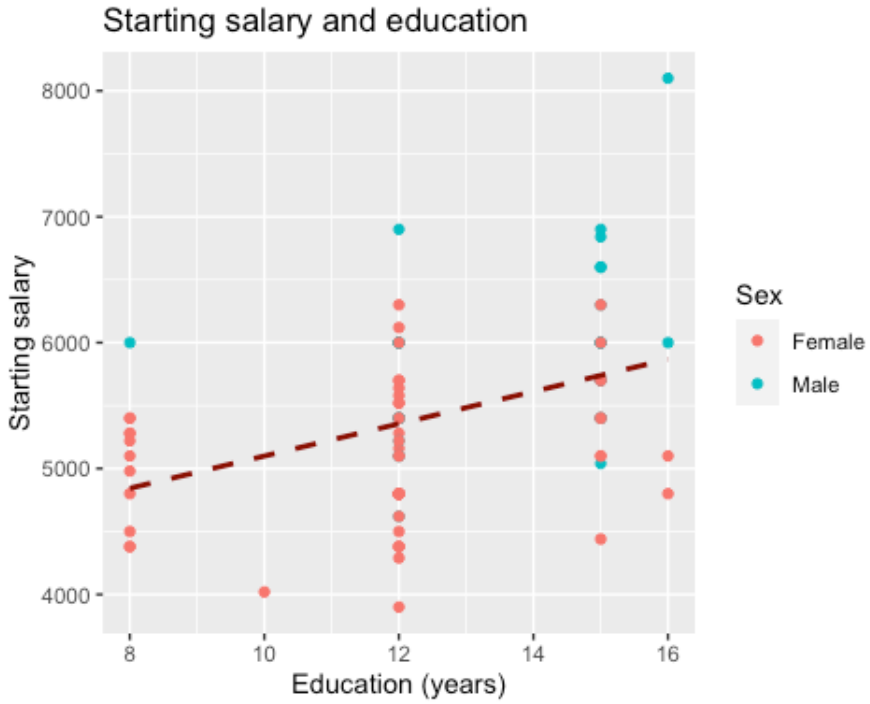
Input variables: Age
Education
Experience
Gender
Seniority

SAMPLES

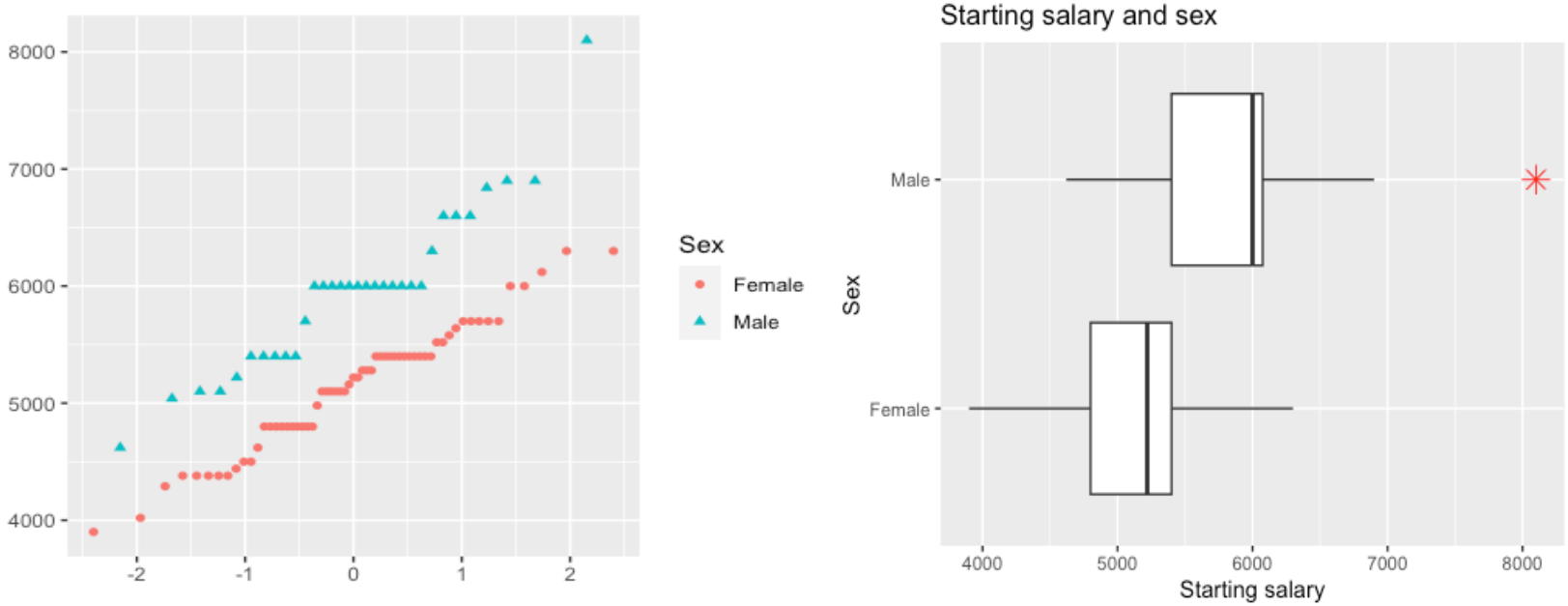
32 men, 61 women

hired between 1965 and 1975

EXPLORATORY GRAPHS

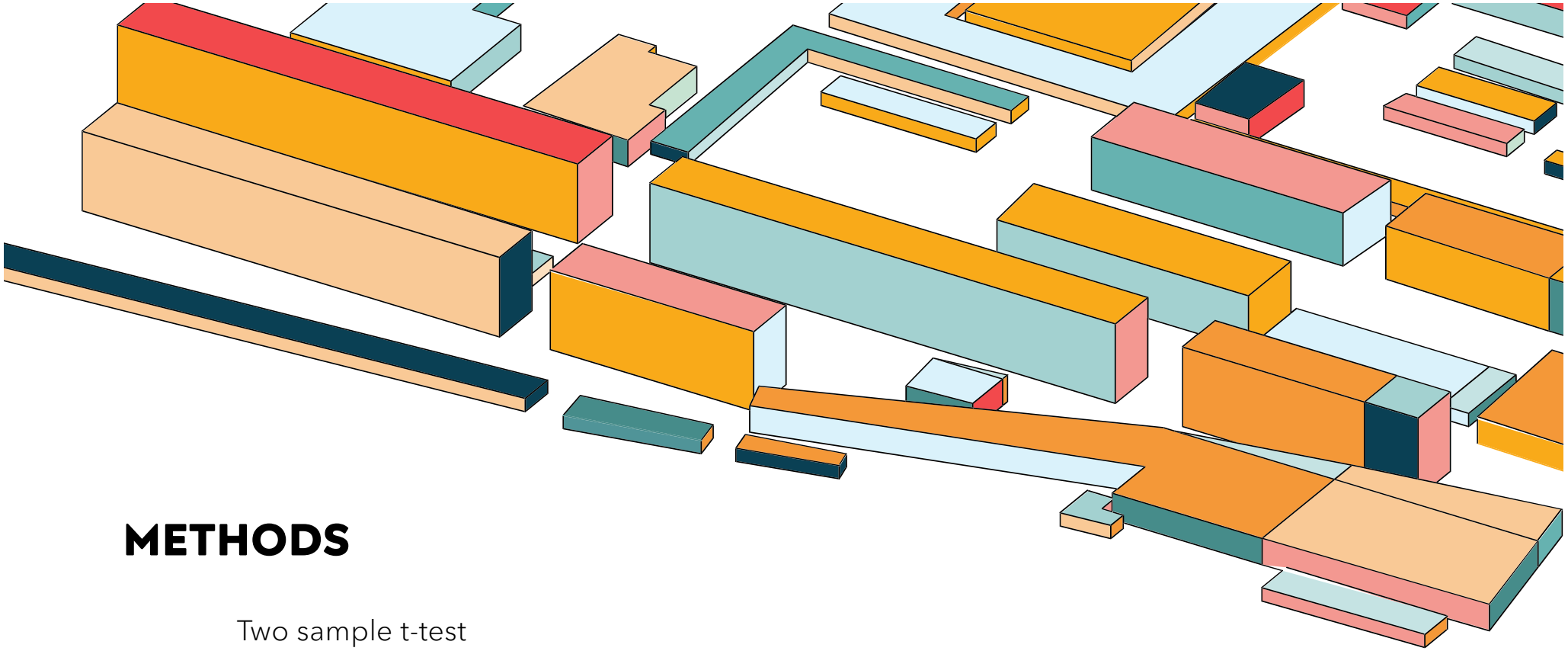


EXPLORATORY GRAPHS



not enough evidence against normality

equal variances



METHODS

Two sample t-test
Regression models controlling for other factors

TWO-SAMPLE T-TEST

- Sample size acceptable
- Not enough evidence against normality
- Equal variances

T-Statistic: -6.2926

Degrees of Freedom: 91

P-Value: 1.076e-08

95% confidence interval:
-1076 -560

Reject the null hypothesis that the true difference in means is equal to 0.

REGRESSION

Goal: model with the smallest number of predictors necessary to provide good predictions.

Forward selection

Successively add predictors, assess based on p-values

Backward elimination

Successively remove predictors, assess based on p-values

Automated – Subsets

Have a computer add, remove, assess predictors; "caret" package step.model "leapBackward"

FORWARD SELECTION – W/O VARIABLE SEX

Linear models	Best: Educ
Non-linear models	No significant effect, not included in the model
Transformations	No effect, not included in the model
Two-variable models	Best: Educ, Senior
Two-way Interactions	No significant effect, not included in the model
Three-variable models	No significant effect, not included in the model
Full model	Sex had significant effect

REGRESSION

Goal: model with the smallest number of predictors necessary to provide good predictions.

Backward elimination

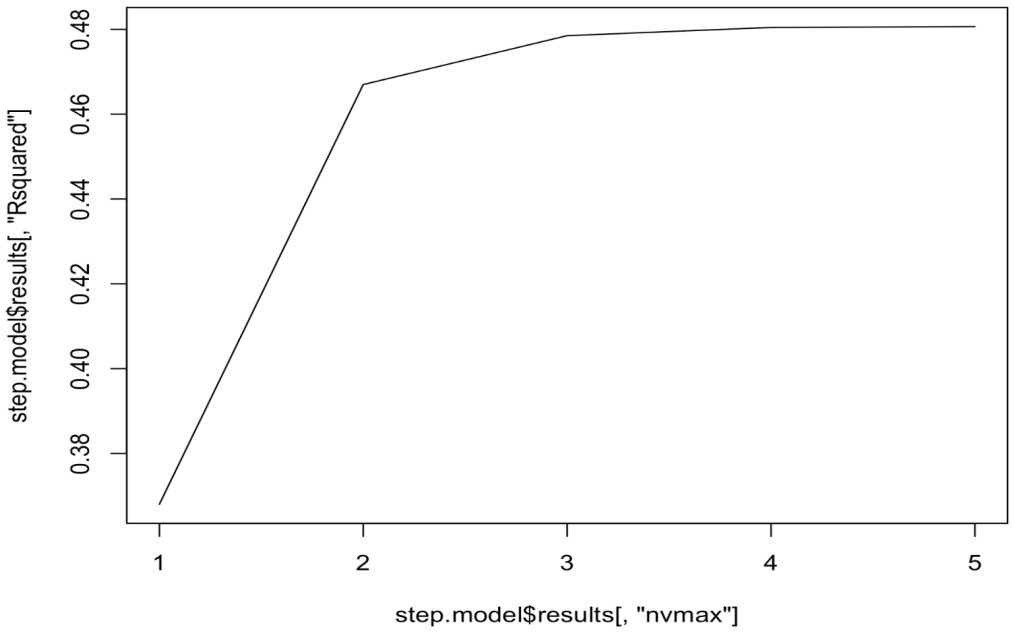
Successively remove predictors, assess based on p-values

Automated – Subsets

Have a computer add, remove, assess predictors; "caret" package step.model "leapBackward"

FINAL MODEL SELECTION

R-squared, # of variables



AIC

MODEL	df	AIC
Full.lm	7	1430.6230
Educ.Senior.Age.Sex.lm	6	1428.8633
Educ.Senior.Sex.lm	5	1432.2413
Senior.Sex.lm	4	1441.3382
Sex.lm	3	1456.3515

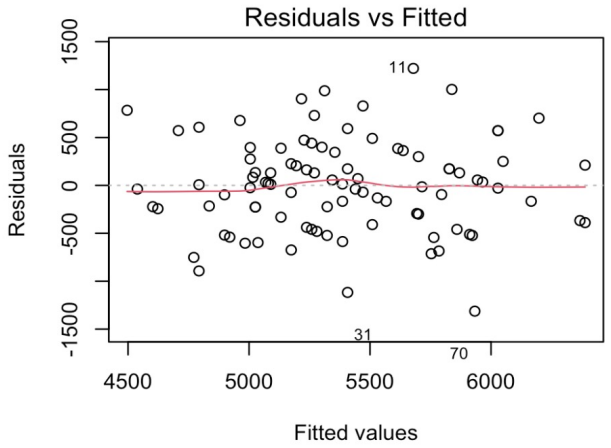
FINAL MODEL

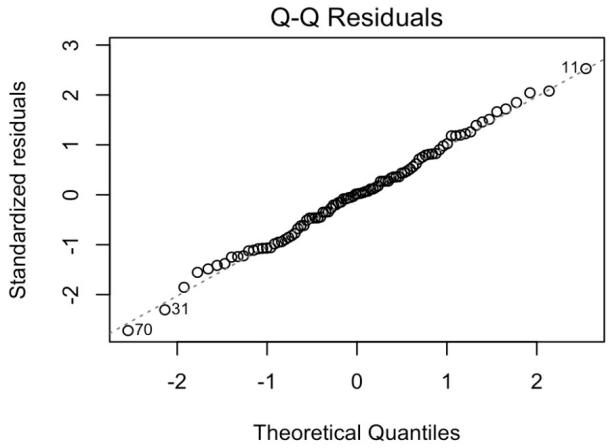
Sex, Seniority, Education

```
1 subsets of each size up to 3
Selection Algorithm: backward
         SexMale Senior Age Educ Exper
1  ( 1 ) "*"     " "    " " " "  " "
2  ( 1 ) "*"     "*"    " " " "  " "
3  ( 1 ) "*"     "*"    " " "*"  " "
```

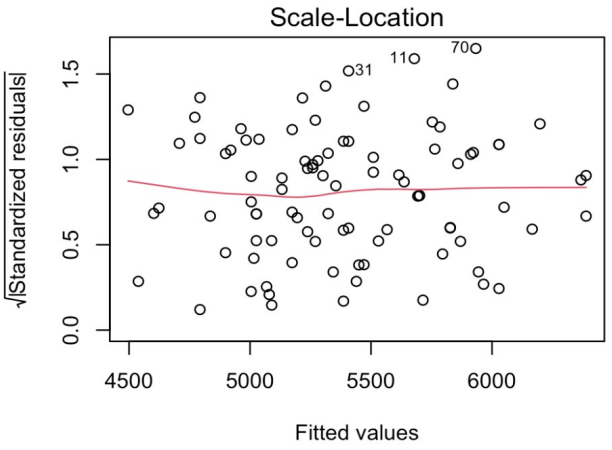
Coefficient table

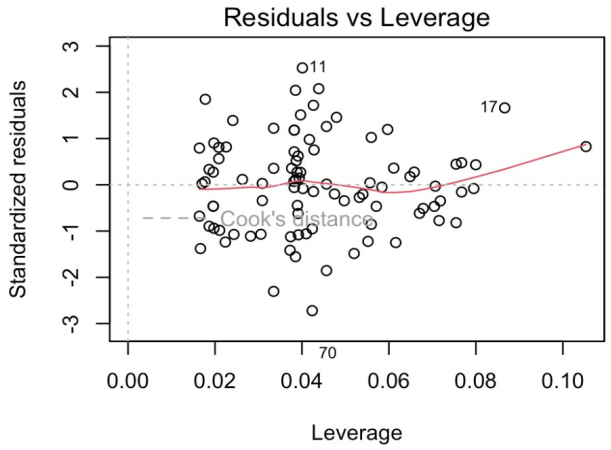
(Intercept)	SexMale	Senior	Educ
6110.79166	737.40468	-24.25706	84.09670





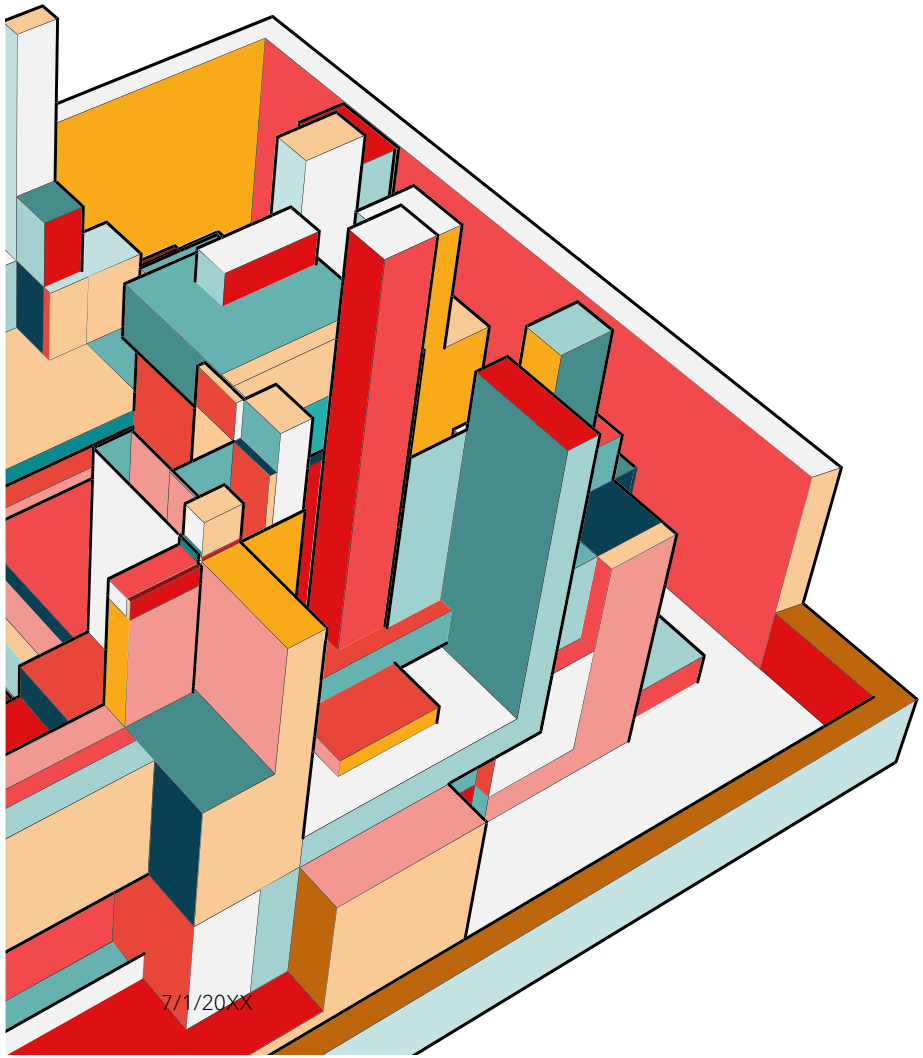
RESIDUAL ANALYSIS





LIMITATIONS

- Observational data, cannot prove causation
- Other variables related to job role not captured
- Qualitative factors also important for legal conclusion



CONCLUSIONS

T-test established that the differences in starting salaries were not due to random chance

Regression established that these differences depended on gender

Conclusion: Statistical evidence suggests gender-based discrimination at the bank, but additional investigation needed for firm legal ruling