Association between policy and fair chance job posting rates

Zofia C. Stanley

2024-10-16

```r
library(ggplot2)

# Set parameter to overwrite plots
overwrite_plots = FALSE

# Read in the data
regression_data <- read.csv("../Controls/regression.csv")
employment_data <- read.csv("../Employment/Indeed/indeed-retail-foodservice-fairchance.csv")
control_data <- read.csv("../Controls/control_indices.csv")
```

Introduction

This analysis aims to explore the relationship between positive policies, collateral consequences, and fair chance hiring outcomes in retail and food service industries. We use scatter plots, linear regression models, and chi-square tests to understand these relationships.

Scatter Plot: Positive Policies vs Collateral Consequences

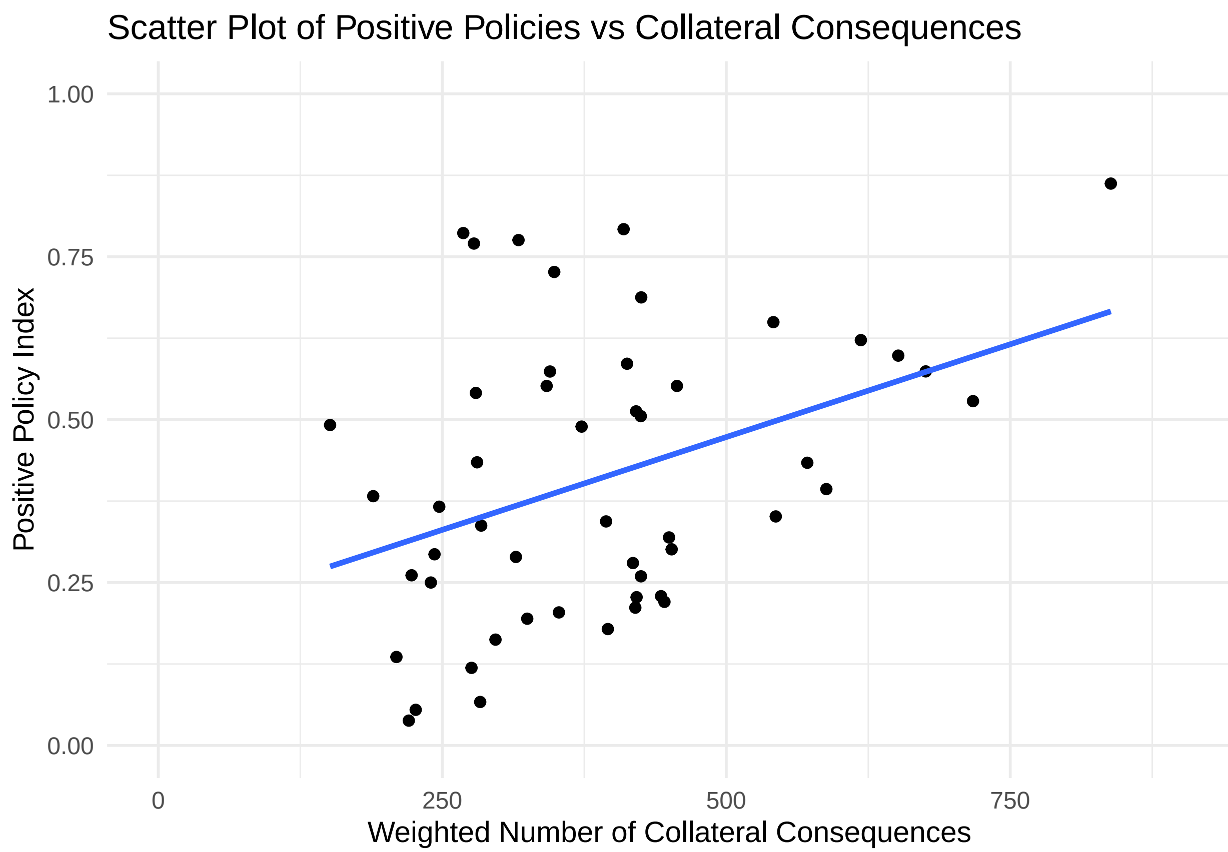
We begin by visualizing the relationship between positive policies and collateral consequences.

```r
# Create a scatter plot of positive policy vs negative policy
plot_pos_vs_neg <- ggplot(regression_data, aes(x = neg_combo_index, y = tot_pos_index)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Scatter Plot of Positive Policies vs Collateral Consequences",
    x = "Weighted Number of Collateral Consequences",
    y = "Positive Policy Index"
  ) +
  xlim(0, 900) +
  ylim(0, 1) +
  theme_minimal()

if (overwrite_plots) {
  pdf("scatter_pos_neg.pdf")
  print(plot_pos_vs_neg)
  dev.off()
}

print(plot_pos_vs_neg)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Interpretation: The scatter plot shows the relationship between the weighted number of collateral consequences and the positive policy index. Visually, we see a strong positive association between the weighted number of collateral consequences and positive policy index.

Scatter Plots: Fair Chance Job Posting Rate vs Policy Indices

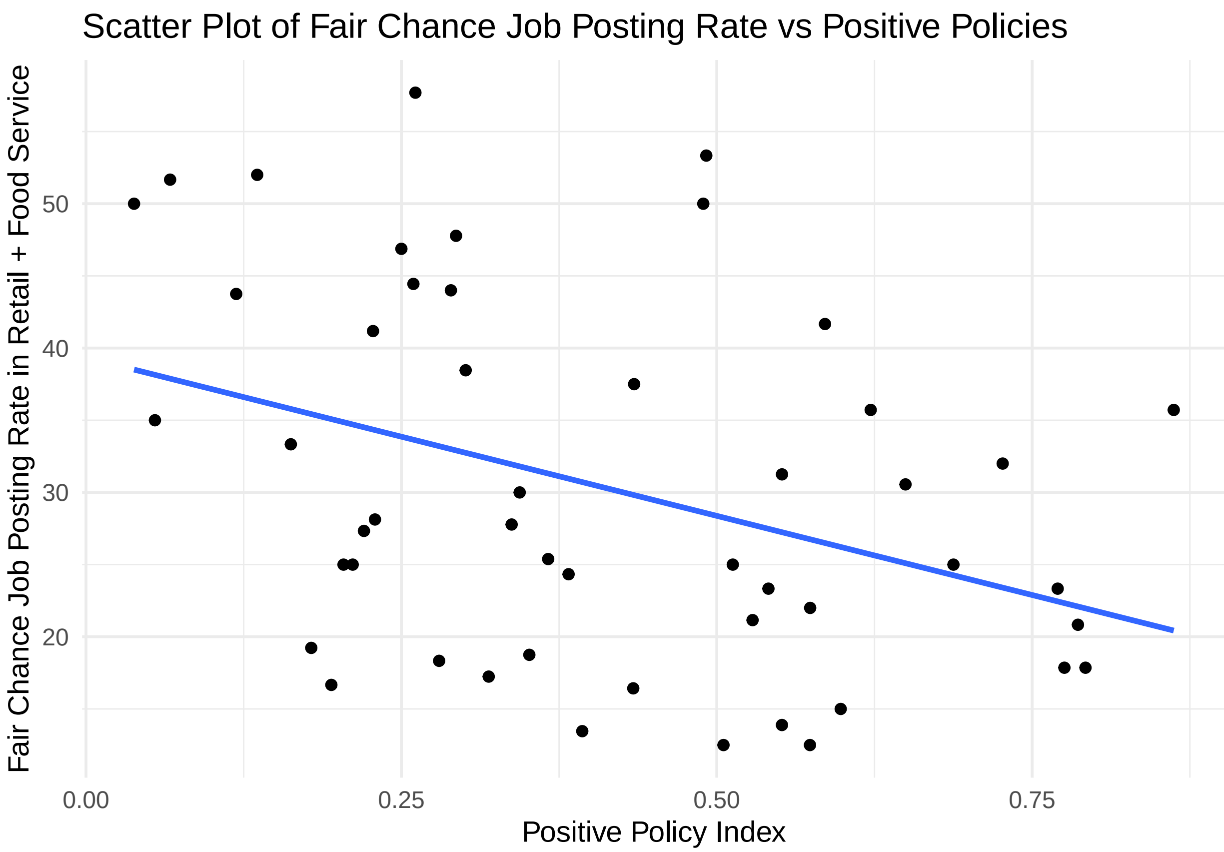
We next look at how fair chance job posting rates relate to both positive and negative policy indices.

```
# Create a scatter plot of positive policy index vs fair chance job posting rate (retail + food service
plot_pos_vs_fairchance_rate <- ggplot(mapping=aes(y = employment_data$fair_chance_rate, x = regression_
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Scatter Plot of Fair Chance Job Posting Rate vs Positive Policies",
    x = "Positive Policy Index",
    y = "Fair Chance Job Posting Rate in Retail + Food Service"
  ) +
  theme_minimal()

if (overwrite_plots) {
  pdf("scatter_pos_fairchance_industry.pdf")
  print(plot_pos_vs_fairchance_rate)
  dev.off()
```

```r
}
print(plot_pos_vs_fairchance_rate)
```

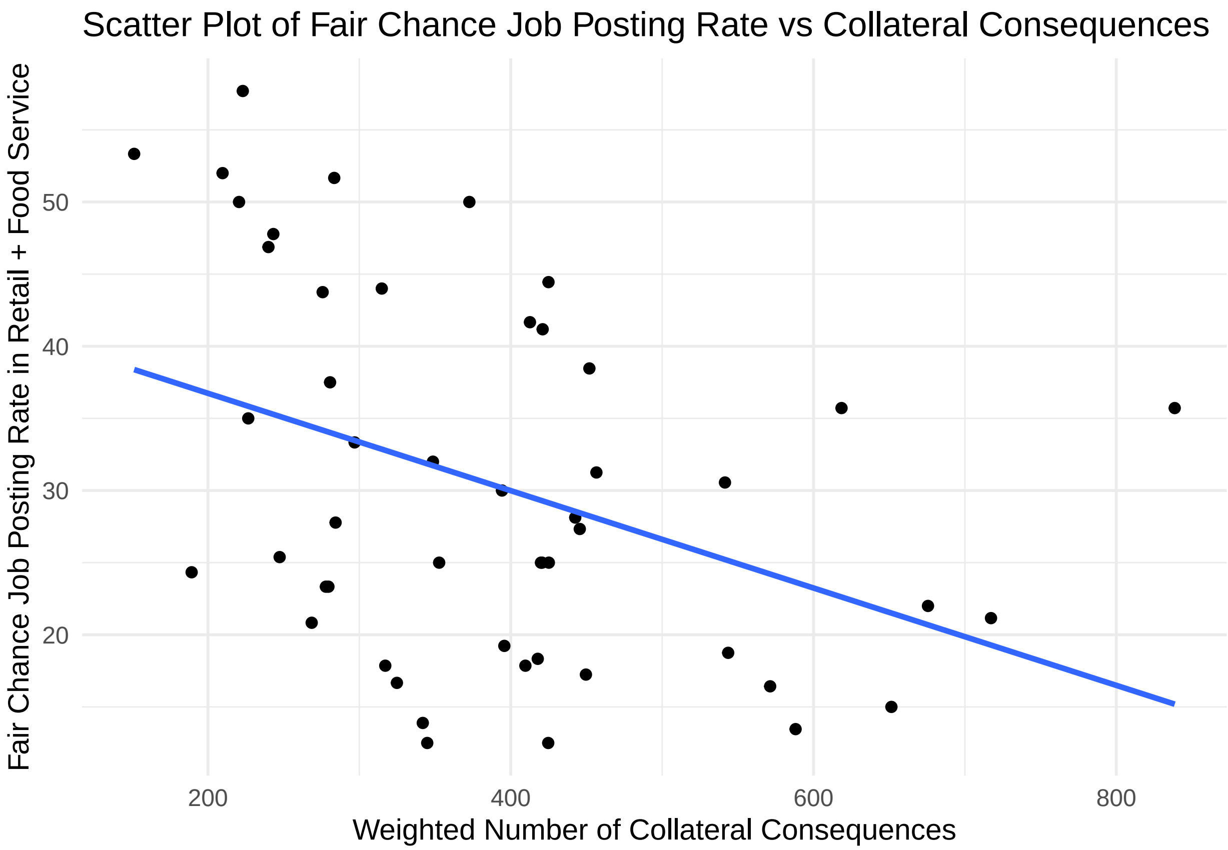
```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
# Create a scatter plot of negative policy index vs fair chance job posting rate in retail and food ser
plot_neg_vs_fairchance_rate <- ggplot(mapping = aes(y = employment_data$fair_chance_rate, x = regressio
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Scatter Plot of Fair Chance Job Posting Rate vs Collateral Consequences",
    x = "Weighted Number of Collateral Consequences",
    y = "Fair Chance Job Posting Rate in Retail + Food Service"
  ) +
  theme_minimal()

if (overwrite_plots) {
  pdf("scatter_neg_fairchance_industry.pdf")
  print(plot_neg_vs_fairchance_rate)
  dev.off()
}
print(plot_neg_vs_fairchance_rate)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Linear Models and Residual Analysis

To control for macroeconomic, socioeconomic, and political factors, we ran several linear models and used residuals to further explore relationships.

```
# Run a linear model for fair_chance_rate vs the three columns of control_data
lm_fairchance_controls <- lm(employment_data$fair_chance_rate ~ control_data$macroeconomy + control_data

# Save the residuals of the model to employment_data$fair_chance_rate_residuals
employment_data$fair_chance_rate_residuals <- residuals(lm_fairchance_controls)

# Print the summary of the linear model
summary(lm_fairchance_controls)
```

```
##
## Call:
## lm(formula = employment_data$fair_chance_rate ~ control_data$macroeconomy +
##     control_data$socioeconomics + control_data$politics)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.427 -11.061  -2.257   9.252  26.579
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   30.345      1.767  17.170   <2e-16 ***
```

```
## control_data$macroeconomy      3.158      2.676   1.180    0.244
## control_data$socioeconomics   -1.518      2.890  -0.525    0.602
## control_data$politics         -2.896      2.056  -1.409    0.166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.5 on 46 degrees of freedom
## Multiple R-squared:  0.0793, Adjusted R-squared:  0.01925
## F-statistic: 1.321 on 3 and 46 DF,  p-value: 0.2791
```

```
# Run a linear model for tot_pos_index vs the three columns of control_data
lm_model_tot_pos <- lm(regression_data$tot_pos_index ~ control_data$macroeconomy + control_data$socioec

# Save the residuals of the model to regression_data$tot_pos_index_residuals
regression_data$tot_pos_index_residuals <- residuals(lm_model_tot_pos)

# Print the summary of the linear model
summary(lm_model_tot_pos)
```

```
##
## Call:
## lm(formula = regression_data$tot_pos_index ~ control_data$macroeconomy +
##     control_data$socioeconomics + control_data$politics)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36547 -0.14020 -0.05289  0.12450  0.43374
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.41029    0.03003  13.661   <2e-16 ***
## control_data$macroeconomy   -0.04600    0.04548  -1.011   0.3172
## control_data$socioeconomics  0.03126    0.04911   0.637   0.5276
## control_data$politics        0.06628    0.03494   1.897   0.0641 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2124 on 46 degrees of freedom
## Multiple R-squared:  0.09779,    Adjusted R-squared:  0.03895
## F-statistic: 1.662 on 3 and 46 DF,  p-value: 0.1883
```

```
# Run a linear model for neg_combo_index vs the three columns of control_data
lm_model_neg_combo <- lm(regression_data$neg_combo_index ~ control_data$macroeconomy + control_data$soci

# Save the residuals of the model to regression_data$neg_combo_index_residuals
regression_data$neg_combo_index_residuals <- residuals(lm_model_neg_combo)

# Print the summary of the linear model
summary(lm_model_neg_combo)
```

```
##
## Call:
## lm(formula = regression_data$neg_combo_index ~ control_data$macroeconomy +
##     control_data$socioeconomics + control_data$politics)
##
## Residuals:
```

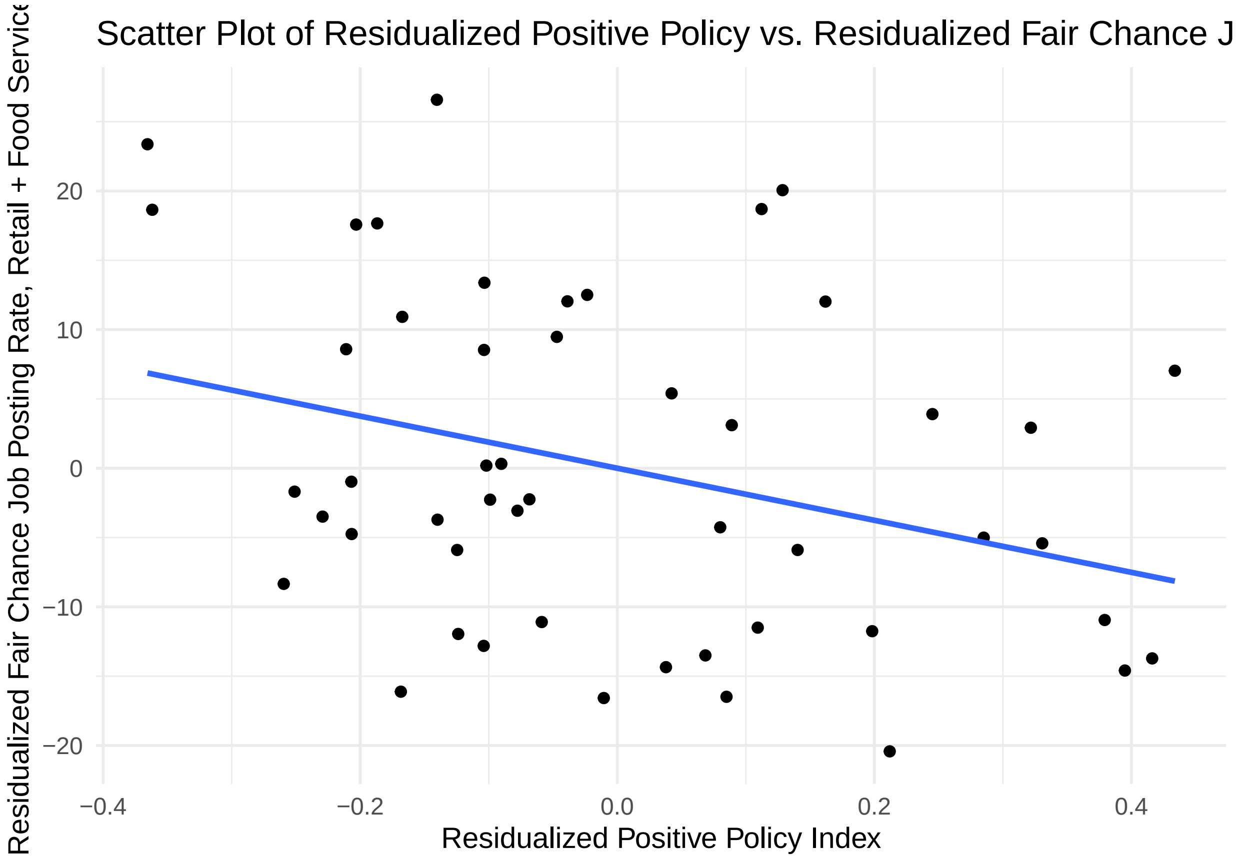
```
##     Min      1Q  Median      3Q     Max
## -206.01  -81.69  -17.02   29.71  413.01
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  389.48      17.31  22.498   <2e-16 ***
## control_data$macroeconomy    -49.71      26.21  -1.896   0.0642 .
## control_data$socioeconomics  -28.35      28.31  -1.002   0.3218
## control_data$politics         50.89      20.14   2.527   0.0150 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 122.4 on 46 degrees of freedom
## Multiple R-squared:  0.3443, Adjusted R-squared:  0.3015
## F-statistic: 8.051 on 3 and 46 DF,  p-value: 0.000204
```

Residualized Analysis

We used residualized values to examine relationships after controlling for external factors.

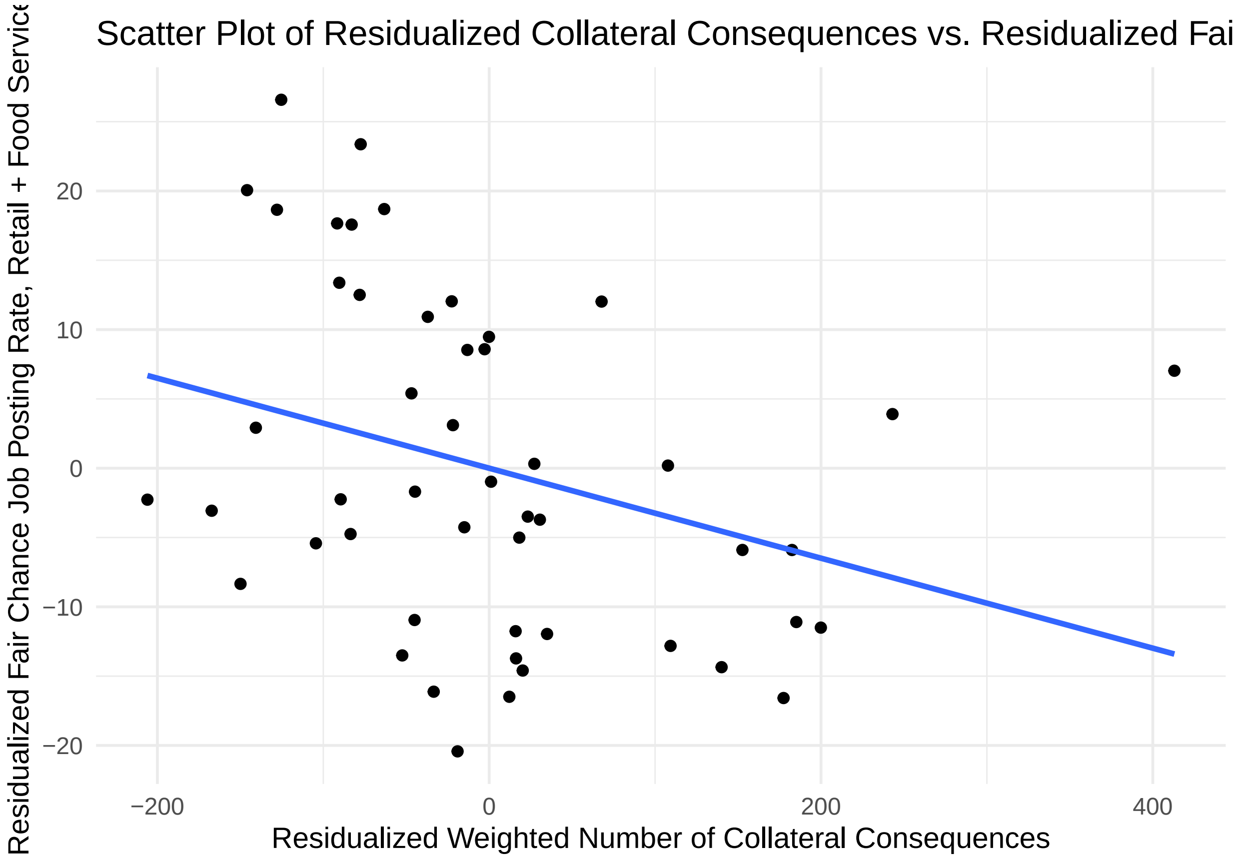
```r
# Scatter plot of residualized positive policy vs. residualized fair chance job posting rate
plot_residual_pos_vs_fairchance <- ggplot(data = data.frame(y = employment_data$fair_chance_rate_residu
  geom_point(aes(x = x, y = y)) +
  geom_smooth(aes(x = x, y = y), method="lm", se=FALSE)+
  labs(
    title = "Scatter Plot of Residualized Positive Policy vs. Residualized Fair Chance Job Posting Rate
    y = "Residualized Fair Chance Job Posting Rate, Retail + Food Service",
    x = "Residualized Positive Policy Index"
  ) +
  theme_minimal()
if (overwrite_plots) {
  pdf("scatter_residual_pos_fairchance.pdf")
  print(plot_residual_pos_vs_fairchance)
  dev.off()
}
print(plot_residual_pos_vs_fairchance)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
# Scatter plot of residualized negative policy vs. residualized fair chance job posting rate
plot_residual_neg_vs_fairchance <- ggplot(data = data.frame(y = employment_data$fair_chance_rate_residua
  geom_point(aes(x = x, y = y)) +
  geom_smooth(aes(x = x, y = y), method="lm", se=FALSE)+
  labs(
    title = "Scatter Plot of Residualized Collateral Consequences vs. Residualized Fair Chance Job Posti
    y = "Residualized Fair Chance Job Posting Rate, Retail + Food Service",
    x = "Residualized Weighted Number of Collateral Consequences"
  ) +
  theme_minimal()
if (overwrite_plots) {
  pdf("scatter_residual_neg_fairchance.pdf")
  print(plot_residual_neg_vs_fairchance)
  dev.off()
}
print(plot_residual_neg_vs_fairchance)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
# Run a linear model for fair_chance_rate vs tot_pos_index, neg_combo_index, and control_data
lm_fair_chance_rate_vs_all_factors <- lm(employment_data$fair_chance_rate ~ regression_data$tot_pos_inde

# Print the summary of the linear model
summary(lm_fair_chance_rate_vs_all_factors)
```

```
##
## Call:
## lm(formula = employment_data$fair_chance_rate ~ regression_data$tot_pos_index +
##     regression_data$neg_combo_index + control_data$macroeconomy +
##     control_data$socioeconomics + control_data$politics)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.367  -8.400  -3.215   9.445  23.472
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     45.86617    6.00686   7.636 1.35e-09 ***
## regression_data$tot_pos_index  -14.35896    8.57884  -1.674    0.101
## regression_data$neg_combo_index -0.02473    0.01488  -1.661    0.104
## control_data$macroeconomy        1.26881    2.61717   0.485    0.630
## control_data$socioeconomics     -1.76981    2.77438  -0.638    0.527
## control_data$politics           -0.68616    2.08887  -0.328    0.744
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 11.75 on 44 degrees of freedom
## Multiple R-squared:  0.2219, Adjusted R-squared:  0.1335
## F-statistic:  2.51 on 5 and 44 DF,  p-value: 0.04392
```

```r
# Run a linear model for residualized fair_chance_rate vs residualized tot_pos_index and neg_combo_index
lm_residual_fair_chance_rate_vs_residual_policies <- lm(employment_data$fair_chance_rate_residuals ~ reg

# Print the summary of the linear model
summary(lm_residual_fair_chance_rate_vs_residual_policies)
```

```
##
## Call:
## lm(formula = employment_data$fair_chance_rate_residuals ~ regression_data$tot_pos_index_residuals +
##     regression_data$neg_combo_index_residuals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.367  -8.400  -3.215   9.445  23.472
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                               1.501e-15  1.607e+00   0.000
## regression_data$tot_pos_index_residuals   -1.436e+01  8.301e+00  -1.730
## regression_data$neg_combo_index_residuals -2.473e-02  1.440e-02  -1.717
##                                           Pr(>|t|)
## (Intercept)                                 1.0000
## regression_data$tot_pos_index_residuals     0.0902 .
## regression_data$neg_combo_index_residuals   0.0926 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.37 on 47 degrees of freedom
## Multiple R-squared:  0.1549, Adjusted R-squared:  0.1189
## F-statistic: 4.308 on 2 and 47 DF,  p-value: 0.01915
```

```r
# Run a linear model for residualized tot_pos_index vs residualized neg_combo_index
lm_residual_positive_policy_vs_residual_negative_combo <- lm(regression_data$tot_pos_index_residuals ~ r

# Print the summary of the linear model
summary(lm_residual_positive_policy_vs_residual_negative_combo)
```

```
##
## Call:
## lm(formula = regression_data$tot_pos_index_residuals ~ regression_data$neg_combo_index_residuals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3237 -0.1579 -0.0321  0.1109  0.4074
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                               -5.008e-18  2.795e-02   0.000
## regression_data$neg_combo_index_residuals  5.386e-04  2.380e-04   2.263
##                                           Pr(>|t|)
## (Intercept)                                 1.0000
## regression_data$neg_combo_index_residuals   0.0282 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1976 on 48 degrees of freedom
## Multiple R-squared:  0.09639,    Adjusted R-squared:  0.07757
## F-statistic:  5.12 on 1 and 48 DF,  p-value: 0.02821
```

Chi-Square Test of Independence

Finally, we perform chi-square tests to understand if having an above-average policy index is associated with above-average employment outcomes.

```r
# Create binary variables for "above or below average" for each measure (Positive Policy Index)
policy_above_avg_pos <- ifelse(regression_data$tot_pos_index_residuals > 0, "Above Avg", "Below Avg")
employment_above_avg <- ifelse(employment_data$fair_chance_rate_residuals > 0, "Above Avg", "Below Avg")

# Create a contingency table based on the two categorical variables (Positive Policy Index)
contingency_table_pos <- table(policy_above_avg_pos, employment_above_avg)

# Print the contingency table (Positive Policy Index)
print(contingency_table_pos)
```

```
##                     employment_above_avg
## policy_above_avg_pos Above Avg Below Avg
##            Above Avg         8        13
##            Below Avg        14        15
```

```r
# Run the chi-square test of independence (Positive Policy Index)
chi_squared_result_pos <- chisq.test(contingency_table_pos)

# Print the test result (Positive Policy Index)
print(chi_squared_result_pos)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  contingency_table_pos
## X-squared = 0.18246, df = 1, p-value = 0.6693
```

Interpretation: The chi-square test of independence for the positive policy index resulted in a p-value of 0.6693, which is much greater than the conventional significance level of 0.05. This means that we fail to reject the null hypothesis, suggesting that there is no statistically significant association between having a positive policy index above average and having a better-than-average employment outcome. In other words, there is no evidence to indicate that states with an above-average positive policy index are more likely to have above-average fair chance job posting rates.

```r
# Create binary variables for "above or below average" for each measure (Positive Policy Index)
policy_above_avg_neg <- ifelse(regression_data$neg_combo_index_residuals > 0, "Above Avg", "Below Avg")
employment_above_avg <- ifelse(employment_data$fair_chance_rate_residuals > 0, "Above Avg", "Below Avg")

# Create a contingency table based on the two categorical variables (Positive Policy Index)
contingency_table_neg <- table(policy_above_avg_neg, employment_above_avg)

# Print the contingency table (Positive Policy Index)
print(contingency_table_neg)
```

```
##                      employment_above_avg
## policy_above_avg_neg Above Avg Below Avg
##            Above Avg         5        16
##            Below Avg        17        12
```

```r
# Run the chi-square test of independence (Positive Policy Index)
chi_squared_result_neg <- chisq.test(contingency_table_neg)

# Print the test result (Positive Policy Index)
print(chi_squared_result_neg)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  contingency_table_neg
## X-squared = 4.6607, df = 1, p-value = 0.03086
```

Interpretation: The chi-square test of independence for the negative combo index yielded a p-value of 0.03086, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis, indicating that there is a statistically significant association between the negative combo policy index being above average and the employment outcome. Specifically, there seems to be an association that suggests states with an above-average negative combo index are more likely to have a below-average employment outcome.

Conclusion

This report provides an exploratory analysis of the relationship between fair chance hiring policies and employment outcomes. We used visualizations, linear regression models, and chi-square tests to assess these relationships, which can provide valuable insights for policy evaluation and improvement.