Analyzing Institutional Outcomes Using Employment Rates and Educational Quality

Group Members: Saurav Shakti Borah, Liana Simopoulos, Ridvan Kanca, Seonkyu Kim, Amrutha Gabbita, Rahul Maurya

Background

Higher education enhances both earning potential and employment rates, with advanced degrees yielding better outcomes. Four-year college graduates generally earn more than those with associate degrees, and higher education is linked to improved health insurance coverage and lower poverty rates (U.S. Bureau of Labor Statistics, 2023). These insights help stakeholders make informed educational choices. The dataset "scorecard" from College Scorecard includes by-college-by-year data on student outcomes in the US. It offers insights into earnings and employment status for students who attended various institutions, providing a comprehensive view of higher education outcomes.

Data Description

A data frame with 48,445 rows and 8 variables

Column Name	Description
unitid	College identifier
inst_name	College or university name
state_abbr	State abbreviation
pred_degree_awarded_ipeds	Predominant degree awarded (1 = less-than-two-year, 2 = two-year, 3 = four-year+)
year	Year of measured outcomes
earnings_med	Median earnings of students who received federal financial aid and began as undergraduates ten years prior
count_not_working	Number of students who are not working (not necessarily unemployed), received federal financial aid, and began as undergraduates ten years prior
count_working	Number of students who are working, received federal financial aid, and began as undergraduates ten years prior

Table 1. Scorecard Data Dictionary

Problem Statement

A college degree is one of the most significant factors affecting the potential median earnings and employment rates. It has been consistently shown in the studies that higher educational attainment results in higher median earnings. For instance, according to a study from 2022, the median weekly earnings for workers without a high school diploma were 20% lower than those with a high school diploma. Moreover, this trend continues with higher educational levels, and higher educational degrees pay more (U.S. Bureau of Labor Statistics, 2023).

Furthermore, higher median earning is not the only benefit of higher education. Higher education also supports a higher employment possibility. For example, the unemployment rate for individuals who have a bachelor's degree or higher is significantly higher than high school diploma owners (National Center for Education Statistics, 2024). We can see from the data that the employment rate among individuals with a bachelor's degree or higher never went below 80% and is nearly reaching 90%, while for those with certificate or associate degrees, it has never reached 80%.

Additionally, the median earnings vary by the type of institution attended like four-year institution graduates typically earn more than those from associate degree graduates. Moreover, higher degrees lead to better health insurance coverage, and they reduce household poverty when compared to lower educational degrees (College Board, 2023). It is also important to note that while certificate programs can increase earning potential in specific fields, this increase is typically less significant than that associated with higher degrees.

As a result, understanding the differences in institutional outcomes can help stakeholders (students, educators, policymakers, employers) to develop informed decisions about their strategies and educational spending. In addition, by recognizing the economic advantages associated with higher education, stakeholders can develop strategies to encourage educational attainment and reduce socioeconomic disparities.

Significance

Higher education significantly enhances individual economic outcomes and broader societal benefits. Studies consistently show that higher educational attainment leads to higher median earnings and lower unemployment rates. For example, the "Smart Money?" study confirms that education increases financial market participation and reduces negative financial events like bankruptcy and foreclosure, indicating that education not only boosts earnings but also improves financial stability and decision-making (Cole et al., 2014). Similarly, the chapter "Beyond Private Gain" emphasizes that the benefits of higher education extend beyond individual gains (Bloom et al., 1970). It highlights that higher education contributes to increased tax revenues, greater productivity, and reduced reliance on government support, emphasizing the critical role of education in driving both personal and economic growth.

These findings highlight the importance of strategic investments in higher education. The human capital theory discussed in "Beyond Private Gain" justifies public investment in education by demonstrating significant returns that benefit society at large(Bloom et al., 1970). These benefits include improved civic engagement, reduced crime rates, and enhanced social cohesion. Such insights highlight the importance of understanding institutional outcomes, helping stakeholders such as students, educators, policymakers, and employers make informed decisions about educational strategies and spending. Effective educational policies that maximize both private and public benefits can lead to a more educated, financially stable, and socially cohesive society.

Practical Impact

Institutions can use this data to assess their own performance and identify areas for improvement, such as enhancing career services or revising academic programs to better align with labor market demands. By analyzing median earnings and employment rates, stakeholders can compare the performance of different institutions, helping to identify which ones are providing better economic outcomes for their graduates. Prospective students can use the data to make informed decisions about where to attend based on potential financial outcomes. Knowing which institutions have higher median earnings and employment rates can guide students towards programs that offer better economic prospects. Students and their families can evaluate the return on investment (ROI) of attending different institutions, considering the cost of education relative to the potential earnings after graduation. Policymakers can use this data to design and implement policies that encourage institutions to improve their performance, such as tying funding to performance metrics like employment rates and median earnings of graduates. Additionally, data on employment outcomes can foster collaborations between educational institutions and industries to develop programs that better prepare students for specific career paths.

Analytical Goal

The primary goal of this analysis is to investigate the impact of education quality, as measured by the work ratio, and the level of education, as indicated by the degree awarded, on median earnings. The study aims to understand how these factors influence earnings outcomes while considering potential confounders and regional differences.

Overview of Analysis

This analysis investigates the impact of education quality (measured by work ratio) and education level (degree awarded) on median earnings using various causal inference methods. The study commences with exploratory data analysis, examining relationships between variables, distributions of earnings, and regional differences, followed by thorough data preparation. Multiple analytical approaches are then employed, including Propensity Score Matching (PSM) using both logistic and probit regression, Difference-in-Differences (DiD) analysis, and calculations of Average Treatment Effect on the Treated (ATT) and Group Average Treatment Effect (GATE). The analysis of education level utilizes additional methods such as Augmented Inverse Probability Weighting (AIPW), Inverse Probability Weighting (IPW), and Partial Linear Model (PLM) estimators. Key findings reveal that institutions with higher work ratios (indicating better education quality) generally show higher median earnings, and awarding higher degrees is associated with significantly higher median earnings. Notably, the study uncovers substantial regional differences in treatment effects, with the Northeast showing positive effects while the Midwest and South often exhibit negative effects. By employing multiple methodologies, the analysis ensures robust results while considering various factors such as regional differences and potential confounders. This comprehensive approach highlights the complex relationship between education quality, degree level, and earnings, underscoring the importance of considering geographical factors in educational policy evaluations.

Detailed Analysis: Education Quality (Work_ratio) on Median

Exploratory Data Analysis

Before starting causal inference analysis, the data was cleaned and analyzed to identify trends and relationships. First, null values were summarized, and rows with null values in "count_not_working" and "count_working" were removed. Missing values in the "earnings_med" variable were imputed using the mean. Feature engineering included creating a "work_ratio" variable, calculated by dividing "count_working" by the total of "count_working" and "count_not_working." States were mapped to US regions for a comprehensive analysis.

Post-cleaning, various visuals were created to identify data trends and relationships. A correlation plot revealed high positive correlations between "earnings_med," "pred_degree_awarded_ipeds," and "work_ratio." A histogram showed a right-skewed distribution of median earnings. A scatterplot indicates that a higher work ratio generally leads to higher median earnings. A boxplot illustrated a stepwise increase in median earnings based on degree type. Bar graphs depicted the count of institutions by region, average median earnings by region, and count of institutions by degree awarded. These visuals, detailed in Appendix A: Exploratory Data Analysis Figures 1-7, clarified the data and variable relationships.

Data Preparation

The analysis starts by performing data preparation, variable transformation, and multicollinearity assessment using variance inflation factors (VIF). It begins by installing and loading required packages, including “car”, then fits a linear regression model with “earnings_med” as the dependent variable and “pred_degree_awarded_ipeds, work_ratio”, and “region” as independent variables. VIF calculations indicate no significant multicollinearity among predictors. The “work_ratio” variable is normalized, and one-hot encoding is applied to the “region” variable. Relevant features are selected, and a binary treatment variable, “work_ratio_binary”, is created based on the median “work_ratio”. Additional packages for matching and propensity score analysis are installed and loaded. The “work_ratio” column is removed, “year” is converted to dummy variables, and treatment and outcome variables are defined. A propensity score formula is created using all covariates except treatment and outcome variables. The head of the “train_data” dataframe, which includes the binary treatment variable, year dummies, and other relevant features, is displayed. This process ensures the data is prepared for causal inference analysis, with no significant multicollinearity and properly formatted variables.

PSM using Logistic Regression

The analysis implements Propensity Score Matching (PSM) using logistic regression and assesses covariate balance post-matching. It begins with package installation and data preparation, including multicollinearity checks via VIF analysis. The “work_ratio” variable is normalized,and the “region” variable undergoes one-hot encoding. A binary treatment variable, “work_ratio_binary”, is created based on the median “work_ratio”. Work ration is serving as a proxy to measure education quality based on the intuition that a higher placement rate suggests a higher-ranked institution with better quality education.

The process continues with further data transformation, including dummy variable creation for the "year" column and defining treatment and outcome variables. A propensity score formula is generated using all relevant covariates. The resulting "train_data" data frame includes the binary treatment variable, year dummies, and other pertinent features.

PSM is then performed using logistic regression with nearest neighbor matching. The matched data is analyzed for covariate balance using standardized mean differences, visualized in a love plot. The balance table indicates minimal differences in covariates post-matching, suggesting good balance. The sample sizes reveal 3510 control units and 15221 treated units after matching, with 11714 control units unmatched.

A linear model estimates the treatment effect for the matched sample. Results show a significant positive effect of the binary treatment on "earnings_med". The intercept (28034.3) represents the estimated median earnings for the control group, while the coefficient for "work_ratio_binary" (12944.0) indicates that the treated group (high work ratio) is associated with an increase of 12944.0 in median earnings, statistically significant at $p < 2e\text{-}16$. The Average Treatment Effect (ATE) for the matched sample is approximately 12944, suggesting that institutions with a high work ratio have, on average, median earnings 12944 higher than those with a low work ratio. Results of this analysis can be found in Appendix B: Education Quality Results in Figures 8 - 11.

PSM using Probit Regression

Next we implemented Propensity Score Matching (PSM) using probit regression, assessing post-matching covariate balance, and estimating the treatment effect. It begins with PSM using probit regression, employing nearest neighbor matching with a probit link function. The process allows unit reuse in matching and discards unmatched units from both treatment and control groups.

Covariate balance after matching is evaluated using standardized mean differences, visualized in a love plot. The balance table indicates minimal differences in covariates post-matching, suggesting good balance. Sample sizes reveal 3510 control units and 15221 treated units after matching, with 11714 control units unmatched.

A linear model fitted to the matched data estimates the treatment effect. Results show a significant positive effect of the binary treatment on "earnings_med". The intercept (28034.3) represents estimated median earnings for the control group, while the coefficient for "work_ratio_binary" (12944.0) indicates that the treated group (high work ratio) is associated with a statistically significant increase of 12944.0 in median earnings ($p < 2e\text{-}16$).

The Average Treatment Effect (ATE) for the matched sample is approximately 12944, suggesting that institutions with a high work ratio have, on average, median earnings 12944 higher than those with a low work ratio. These results indicate that the probit regression-based matching process successfully balanced covariates between treatment and control groups, and the treatment (high work ratio) significantly positively affects median earnings. Appendix B: Education Quality Results in Figures 12 - 15.

Difference-in-Difference Analysis

Then a Difference-in-Differences (DiD) analysis was implemented to evaluate the effect of a high work ratio on median earnings. It begins by creating a binary indicator for pre- and post-2012 periods, then groups and aggregates data by “unitid” and this indicator. Median values for “work_ratio” and “earnings_med” are calculated, along with maximum values for region and degree level variables. A binary work ratio column is then created to classify high and low work ratios.

The data is filtered to include only units with both pre- and post-2012 data. Treatment units are defined as those transitioning from low to high work ratio after 2012, while control units maintain a low work ratio throughout. The final dataset combines these units, and a treatment group indicator is added.

A DiD model is fitted to estimate the treatment's impact. The key coefficient of interest, representing the interaction between treatment group and post-treatment period (“time_treatment”), is 513.39 but not statistically significant ($p = 0.309$). This suggests no significant additional effect of the treatment on median earnings after 2012. Other coefficients show expected associations between covariates and median earnings.

To validate the DiD approach, a parallel trends check is conducted by plotting pre-2012 median earnings for both groups. The plot demonstrates similar trends, supporting the parallel trends assumption. While the treatment group shows higher median earnings overall, the lack of statistical significance in the treatment effect post-2012 suggests other factors may be influencing the observed increase in median earnings for the treatment group. Results can be found in Appendix B: Education Quality Results Figures 16 and 17.

ATT and GATE Calculations

Lastly, the Average Treatment Effect on the Treated (ATT) and the Group Average Treatment Effect (GATE) by region, then visualizes the GATE estimates. It begins by extracting the ATT from the Difference-in-Differences (DiD) model, summing the coefficients for “after_2012” and “time_treatment”. The resulting ATT of approximately 506 suggests that the treatment (high work ratio) increases median earnings by an average of 506 for treated units.

Next, the GATE is calculated by fitting a linear model that interacts “work_ratio_binary” with regional dummy variables. The estimates of these interaction terms, representing region-specific treatment effects, are extracted. Results show significant regional variation: the Northeast exhibits a positive treatment effect of approximately 1329, indicating an increase in median earnings due to the treatment. Conversely, the Midwest and South show negative effects of about -2497 and -2924 respectively, suggesting the treatment decreases median earnings in these regions.

A bar plot is then created to visualize these GATE estimates, clearly illustrating the regional differences in treatment effects. The Northeast's positive effect contrasts sharply with the negative effects observed in the Midwest and South. This analysis highlights the significant regional disparities in how a high work ratio impacts median earnings, with positive outcomes in the Northeast but negative results in the Midwest and South suggesting that degree ROI isn't consistent regionally.

Detailed Analysis: Education Level (Degree_level) on Median Earnings

PSM with ATE Calculations

The analysis starts with conducting a Propensity Score Matching (PSM) analysis using logistic regression to estimate how the predominant degree awarded affects median earnings. It begins by loading necessary libraries and transforming the "pred_degree_awarded_ipeds" variable into a binary column, with degrees 1 and 2 recorded as 0, and others as 1. The `year` column is converted to dummy variables. Treatment and outcome variables are defined as "pred_degree_awarded_ipeds_binary" and "earnings_med", respectively. A logistic regression model is then employed for PSM, matching treated and control units.

Post-matching covariate balance is assessed using standardized mean differences, visualized in a love plot. Results indicate well-balanced covariates after matching, with adjusted mean differences near zero for all covariates, suggesting successful matching. The sample sizes reveal 3128 control units and 10978 treated units after matching, with 16334 control units remaining unmatched.

A linear model summary estimates the treatment effect on median earnings. The intercept (33845.5) represents estimated median earnings for institutions awarding lower degrees. The coefficient for "pred_degree_awarded_ipeds_binary" (9470.0) indicates that institutions awarding higher degrees have median earnings averaging 9470 higher than those awarding lower degrees, a statistically significant result ($p < 2e-16$). These results can be seen in Appendix C: Education Quality Results in Figures 20 - 23.

The Average Treatment Effect (ATE) for the matched sample is approximately 9470, indicating a substantial positive effect of awarding higher degrees on median earnings. Next, the analysis estimates the Average Treatment Effect (ATE) using a regression-based approach, confirming that institutions awarding higher degrees have median earnings that are approximately 9469.97 higher than those awarding lower degrees. This result is consistent with previous analyses, indicating a significant positive effect of awarding higher degrees on median earnings. The installation of additional packages suggests preparation for further advanced analyses. In summary, the PSM process effectively balanced covariates, and the estimated treatment effect shows a significant increase in median earnings for institutions awarding higher degrees.

AIPW Estimator

Next we implemented an augmented inverse probability weighted (AIPW) estimation to calculate the Average Treatment Effect (ATE). It begins by fitting a propensity score model using the SuperLearner package, incorporating generalized linear models, stepwise regression, and elastic net regression. Predicted propensity scores are added to the dataset, with extreme scores trimmed to avoid extreme weights. Weights are then recalculated and stabilized to reduce variance.

An outcome regression model is fitted using the same algorithms, and predicted outcomes are computed. The augmentation component, combining weighted residuals from the outcome model with stabilized weights, is calculated. The ATE is then derived as the mean of this augmentation component. Propensity scores ranged from 0 to nearly 1, with a median of approximately 0.21, indicating variability in treatment probability across units. Stabilized weights had a mean of 8.96, with a maximum of 957, showing some high influence points. Augmentation values varied widely, reflecting differences between observed and predicted outcomes. Results can be seen in Figure 24 in Appendix C: Education Quality Results.

The AIPW method yielded an ATE of approximately 5405, suggesting that institutions awarding higher degrees have median earnings averaging 5405 higher than those awarding lower degrees. This result, incorporating both propensity score weighting and outcome regression adjustments, provides a robust estimate of the treatment effect.

IPW Estimator

The code estimates the Average Treatment Effect (ATE) using Inverse Probability Weighting (IPW) with R's “survey” package. It begins by loading necessary libraries: “survey”, “SuperLearner”, “dplyr”, and “fastDummie”`. A survey design object is created using “svydesign:, incorporating stabilized weights from the dataset.

A generalized linear model (GLM) is then fitted using “svyglm”, with “earnings_med” as the outcome variable and “pred_degree_awarded_ipeds_binary” as the treatment variable. The ATE is represented by the treatment variable's coefficient, extracted from the model summary.

The resulting ATE of approximately 6193.68 suggests that institutions awarding higher degrees have median earnings averaging 6193.68 higher than those awarding lower degrees. This significant positive effect corroborates previous analyses, confirming that awarding higher degrees substantially impacts median earnings.

PLM Estimator

The code estimates the Average Treatment Effect (ATE) using a nonparametric regression approach. It begins by loading the required library and constructing a formula for the nonparametric regression model, specifying “earnings_med” as the outcome, “pred_degree_awarded_ipeds_binary” as the treatment, and including relevant covariates. All variables are converted to numeric format.

The “npregbw” function estimates the bandwidth for the nonparametric regression, which is then used by `npreg` to fit the model. Outcomes are predicted using this fitted model, and the ATE is calculated as the mean of these predictions. The result is presented as "ATE (PLM): (value)," indicating the average effect of awarding higher degrees on median earnings.

Timing information is provided, showing the computational effort for model fitting using multistart optimization. This nonparametric approach offers a flexible and robust estimate of the treatment effect, accommodating complex relationships between variables without assuming a specific parametric form.

GATE Analysis

The code estimates and visualizes the Group Average Treatment Effect (GATE) by region using a linear model. A formula is constructed for the linear model, with "earnings_med" as the outcome, "pred_degree_awarded_ipeds_binary" as the treatment, interaction terms with regional dummy variables, and `work_ratio` as a covariate.

After fitting the linear model, GATE estimates for each region are extracted from the summary. Results show a positive treatment effect in the Northeast (estimate ≈ 656), indicating increased median earnings. Conversely, the Midwest and South exhibit negative effects (estimates ≈ -382 and -653 respectively), suggesting decreased median earnings.

A bar plot is created to visually represent these GATE estimates as seen in Appendix C: Education Level Results in Figure 25, clearly illustrating the positive effect in the Northeast and negative effects in the Midwest and South. This analysis highlights significant regional disparities in how awarding higher degrees impacts median earnings, emphasizing the importance of considering geographical factors when evaluating educational policies.

Results

Discussion

This study offers valuable insights into the impact of education quality and level on median earnings, with significant implications for educational policies. The findings consistently demonstrate that institutions with higher work ratios (a proxy for education quality) are associated with substantially higher median earnings, with an Average Treatment Effect (ATE) of about $12,944. Similarly, awarding higher degrees is linked to increased median earnings, with ATE estimates ranging from $5,405 to $9,470 across different methodologies. However, the study also reveals significant regional disparities, with the Northeast generally showing positive effects, while the Midwest and South often exhibit negative effects. This regional variation suggests the need for tailored educational policies that address specific local challenges and opportunities. The Difference-in-Differences analysis, showing no significant additional effect of high work ratio on median earnings after 2012, further highlights the complexity of these relationships and the influence of evolving economic factors. These findings have several policy implications, including the need for quality-focused initiatives, improved higher education access, regionally tailored strategies, and a holistic approach to educational policy. The study supports policies that prioritize improving education quality, increasing access to higher education, and aligning educational programs with regional job market needs. It also emphasizes the importance of continuous policy evaluation and adjustment to ensure relevance and effectiveness in changing economic landscapes. In conclusion, while the study strongly supports the economic value of both education quality and higher degree attainment, it also underscores the complexity of these relationships and the need for nuanced, data-driven educational policies that consider regional variations and evolving economic contexts.

Limitations

While the study provides valuable insights into the relationship between higher education and economic outcomes, several limitations must be acknowledged that may impact the interpretation and generalizability of the findings. The dataset primarily includes information on students who have received federal financial aid, potentially limiting its representation of the entire student population. This focus may skew the analysis towards outcomes specific to students who qualify for federal aid, potentially overlooking the experiences of students from different socioeconomic backgrounds. The dataset captures outcomes ten years after students begin their educational journey, offering valuable insights into long-term economic impacts, but it may not fully capture shorter-term outcomes or consider changes in economic conditions over time.

Due to the observational nature of the data, establishing causal relationships between educational attainment and economic outcomes presents challenges. While the analysis identifies correlations, it cannot definitively attribute changes in economic status solely to educational attainment, as other factors such as economic fluctuations and individual career choices may also influence outcomes. The findings primarily pertain to the U.S. higher education system and may not be directly applicable to other countries or educational contexts with different structures and economic environments. Factors such as educational policies, labor market dynamics, and cultural norms can significantly influence how educational attainment translates into economic outcomes across different regions. Furthermore, the dataset lacks comprehensive information on critical socioeconomic variables such as family income, parental education levels, and geographic location. These factors play a significant role in shaping educational opportunities and outcomes, and their absence limits the study's ability to account for broader socioeconomic influences on educational attainment and subsequent economic success.

Future Direction

Future analysis should integrate comprehensive socioeconomic data to better understand how education quality and degree level impact median earnings. Variables such as family income, parental education levels, and geographic location would provide a nuanced view of the interplay between socioeconomic factors and educational outcomes, revealing how different backgrounds influence education's effectiveness in improving economic status. Extending the temporal scope to include both short-term and long-term outcomes could offer insights into the immediate and evolving economic impacts of educational attainment. Tracking economic outcomes over time would allow for an assessment of how post-education earnings progress and how economic conditions influence these trends.

Applying advanced causal inference techniques, such as instrumental variables or regression discontinuity designs, could strengthen causal claims regarding the impact of education on economic outcomes. These methods would better isolate the effect of educational attainment from other confounding factors. Expanding the analysis to include comparisons across different countries or regions could highlight how varying educational policies, labor market dynamics, and cultural norms influence the relationship between education and economic success. This comparative approach would provide valuable insights into the contextual factors shaping educational outcomes globally. In summary, future research should integrate socioeconomic data, extend the temporal scope, employ advanced causal inference methods, conduct cross-regional comparisons, and incorporate qualitative insights to build a comprehensive and actionable understanding of the impact of education on economic outcomes.

References

Bloom, D. E., Hartley, M., & Rosovsky, H. (1970b, January 1). *Beyond private gain: The public benefits of Higher Education*. SpringerLink. https://link.springer.com/chapter/10.1007/978-1-4020-4012-2_15

Cole, S., Paulson, A., & Shastry, G. K. (2014, February 26). *Smart money? the effect of education on financial outcomes*. OUP Academic. https://academic.oup.com/rfs/article-abstract/27/7/2022/1578758?redirectedFrom=fulltext

College Board. (2023). *Education pays, 2023.* https://research.collegeboard.org/media/pdf/education-pays-2023.pdf

Huntington-Klein, N. (2022, October 12). *Example data sets for causal inference textbooks*. Package ‘causaldata.’ https://cran.r-project.org/web/packages/causaldata/causaldata.pdf

National Center for Education Statistics. (2024). *Employment and Unemployment Rates by Educational Attainment.* https://nces.ed.gov/programs/coe/indicator/cbc

U.S. Bureau of Labor Statistics. (2023). *Education pays, 2022.* https://www.bls.gov/careeroutlook/2023/data-on-display/education-pays.htm

Appendix A: Exploratory Data Analysis

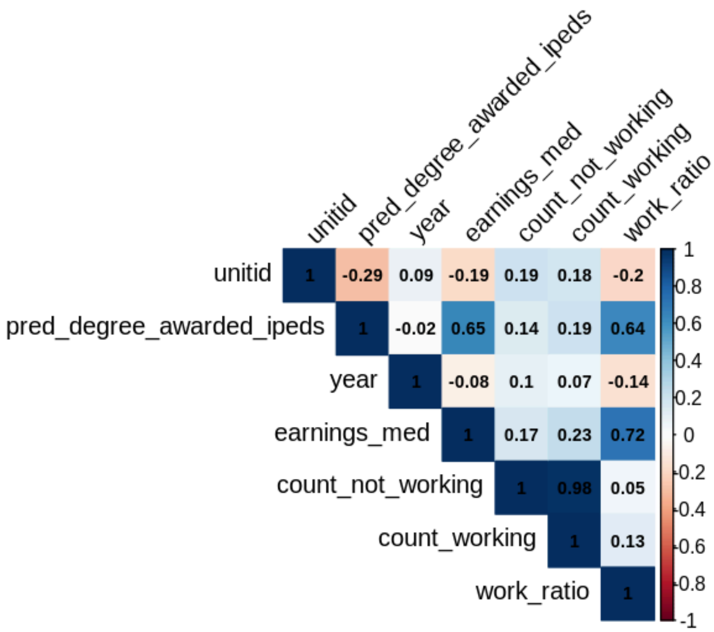

Figure 1. Correlation Plot of Scorecard Variables

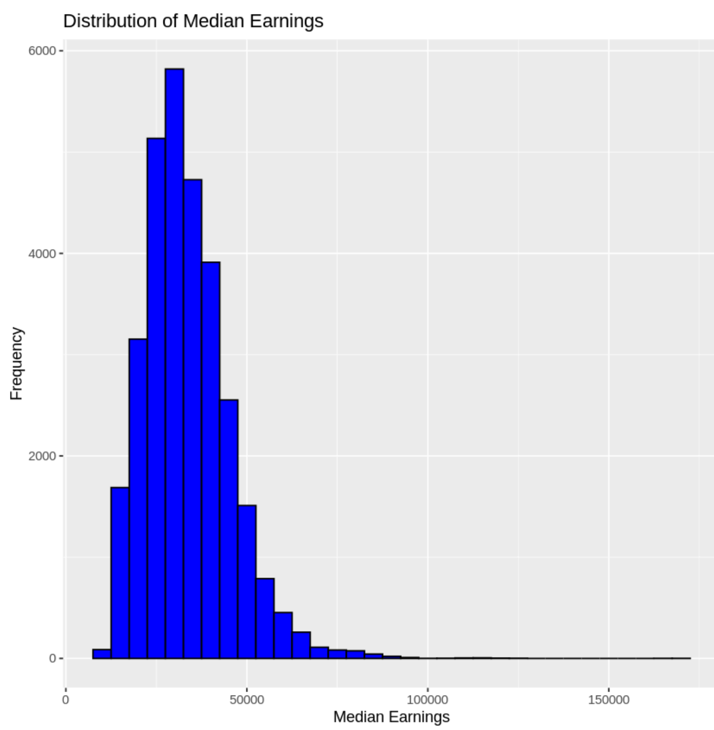

Figure 2. Distribution of Median Earning from Scorecard Dataset

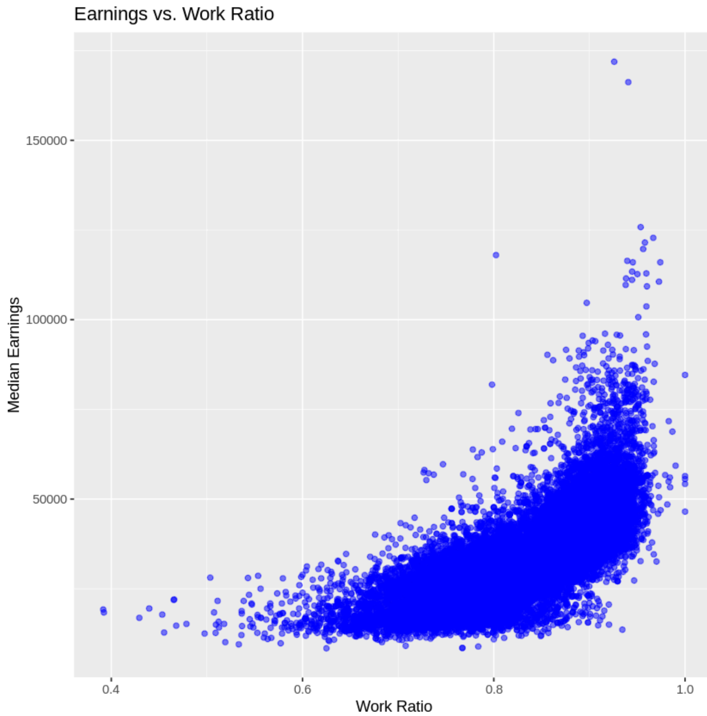

Figure 3. Scatterplot of Median Earning vs. Work Ratio

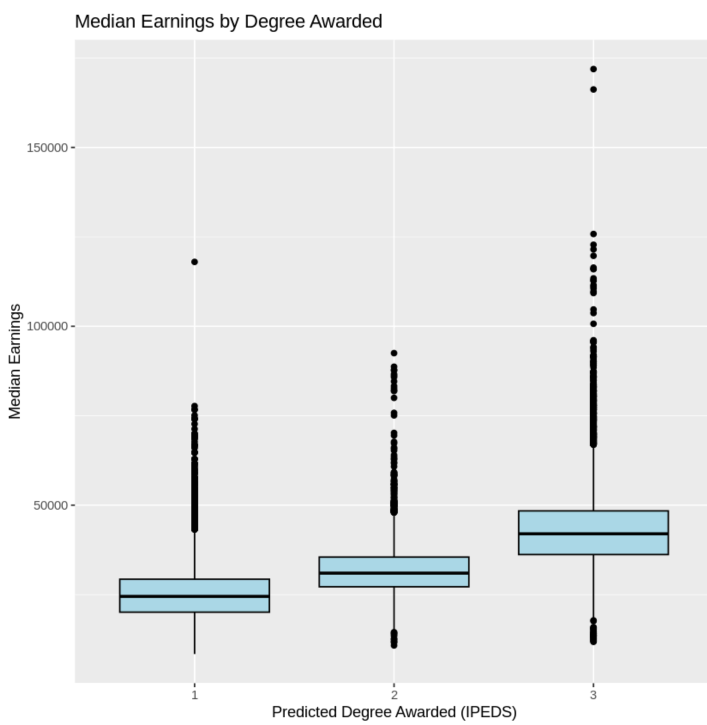

Figure 4. Boxplot of Median Earnings by the Degree Awarded

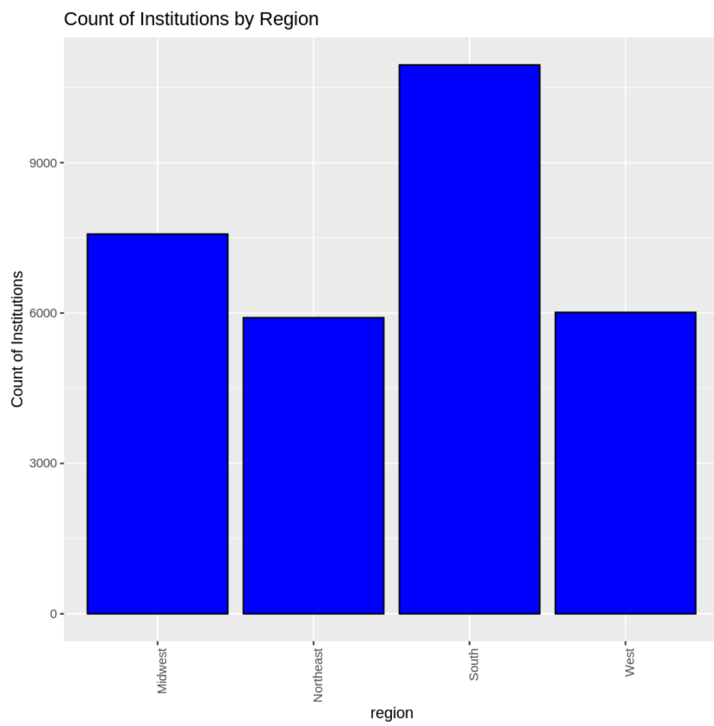

Figure 5. Bar Graph Showing Count of Institutions by the Region

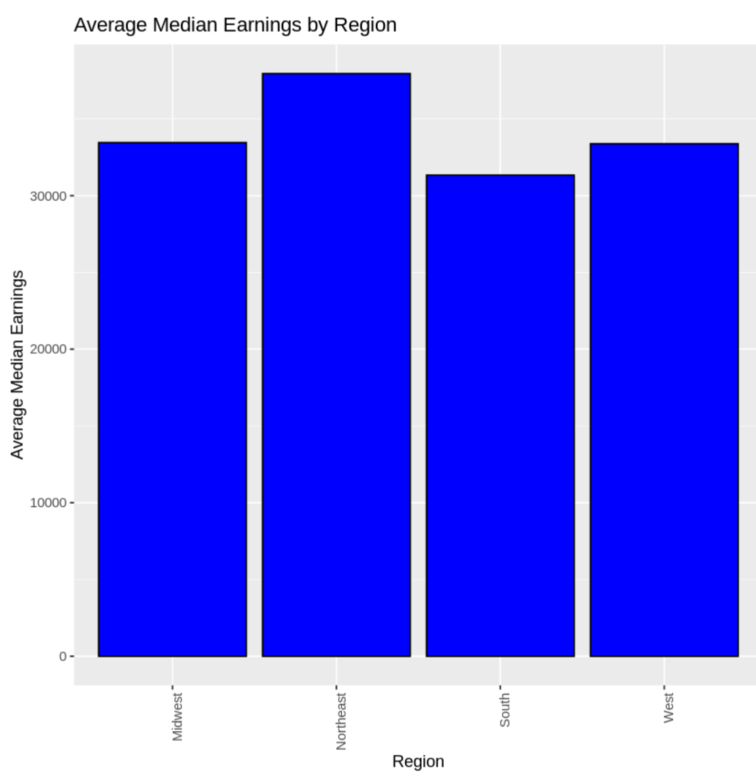

Figure 6. Bar Graph Showing Average Median Earnings by Region

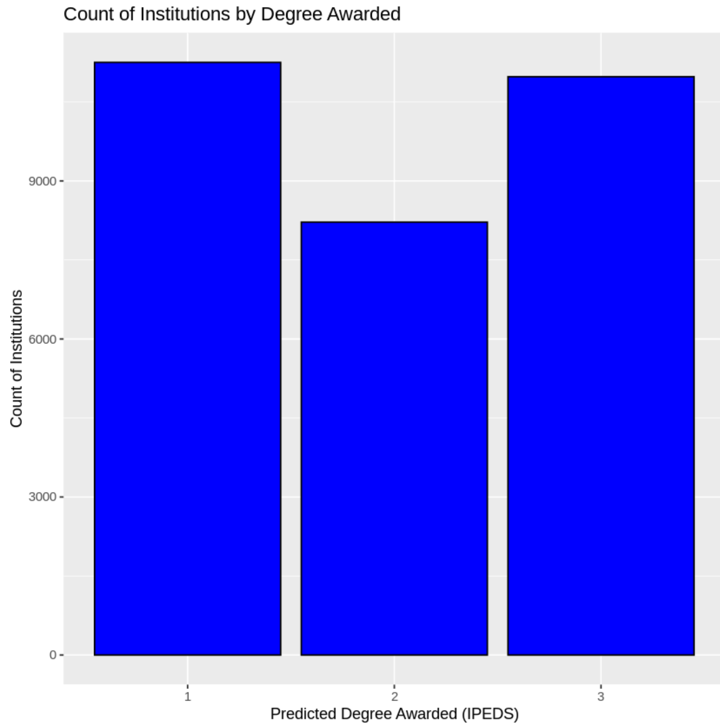

Figure 7. Bar Graph showing the Count of Institutions by the Predicted Degree Awarded

Appendix B: Education Quality Analysis Results

```
Balance Measures
                                 Type Diff.Adj
distance                     Distance        0
pred_degree_awarded_ipeds     Contin.        0
regionNortheast                Binary        0
regionMidwest                  Binary        0
regionSouth                    Binary        0
regionWest                     Binary        0
year_2007                      Binary        0
year_2009                      Binary        0
year_2011                      Binary        0
year_2012                      Binary        0
year_2013                      Binary        0
year_2014                      Binary       -0
```

Figure 8. Summary of Balance Measures from PSM using Logistic Regression

```
Sample sizes
                      Control Treated
All                   15224.    15221
Matched (ESS)           826.06  15221
Matched (Unweighted)   3510.    15221
Unmatched             11714.        0
```

Figure 9. Summary of Sample Sizes from PSM using Logistic Regression

```
Call:
lm(formula = as.formula(paste(outcome, "~", treatment)), data = matched_data_logistic)

Residuals:
   Min     1Q Median     3Q    Max
-29178  -6478  -1278   5122 130922

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        28034.3      175.9  159.34   <2e-16 ***
work_ratio_binary  12944.0      195.2   66.32   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 10. Model Summary from PSM using Logistic Regression

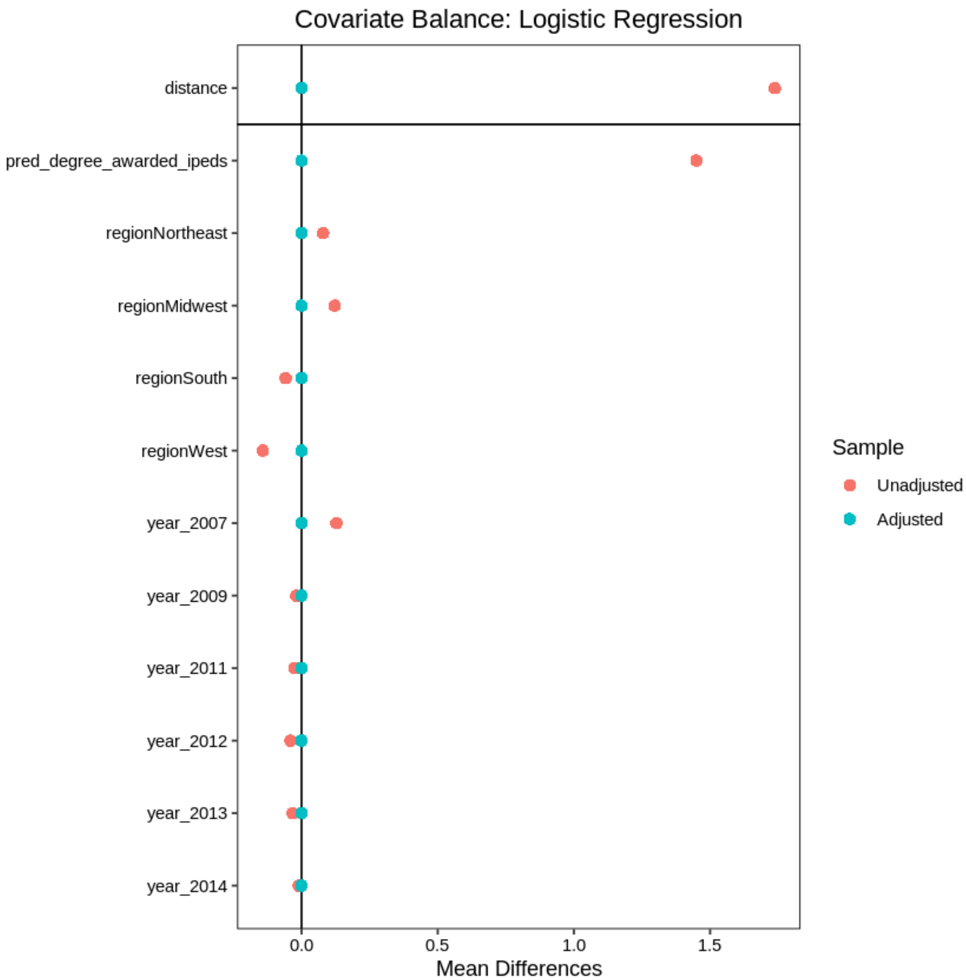

Figure 11. Love Plot from PSM using Logistic Regression

```
Balance Measures
                               Type Diff.Adj
distance                   Distance        0
pred_degree_awarded_ipeds   Contin.        0
regionNortheast              Binary        0
regionMidwest                Binary        0
regionSouth                  Binary        0
regionWest                   Binary        0
year_2007                    Binary        0
year_2009                    Binary        0
year_2011                    Binary        0
year_2012                    Binary        0
year_2013                    Binary        0
year_2014                    Binary       -0
```

Figure 12. Summary of Balance Measure for PSM using Probit Regression

```
Sample sizes
                        Control Treated
All                     15224.    15221
Matched (ESS)             826.06  15221
Matched (Unweighted)     3510.    15221
Unmatched               11714.        0
```

Figure 13. Summary of Sample Sizes for PSM using Probit Regression

```
Call:
lm(formula = as.formula(paste(outcome, "~", treatment)), data = matched_data_probit)

Residuals:
   Min     1Q Median     3Q    Max 
-29178  -6478  -1278   5122 130922 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        28034.3      175.9  159.34   <2e-16 ***
work_ratio_binary  12944.0      195.2   66.32   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 14. Model Summary from PSM using Probit Regression

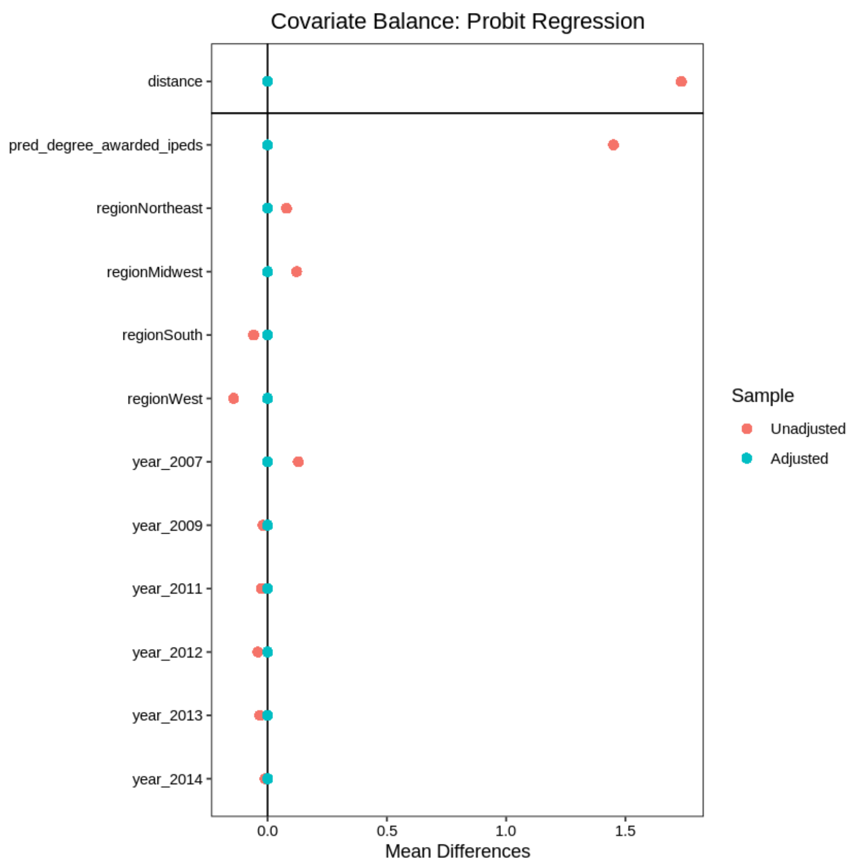

Figure 15. Love Plot from using PSM with Probit Regression

```
Call:
lm(formula = earnings_med ~ after_2012 + treatment_group + degree_level +
    time_treatment + regionNortheast + regionMidwest + regionSouth +
    regionWest, data = final_data)

Residuals:
   Min      1Q Median      3Q     Max
-18125  -3736    -44    3386   42219

Coefficients: (1 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     21699.71     255.41  84.960  < 2e-16 ***
after_2012         -7.61     163.91  -0.046    0.963
treatment_group  2546.58     360.08   7.072 1.73e-12 ***
degree_level     3944.28     126.00  31.302  < 2e-16 ***
time_treatment    513.39     505.11   1.016    0.309
regionNortheast -1161.54     250.04  -4.645 3.48e-06 ***
regionMidwest   -3078.23     237.46 -12.963  < 2e-16 ***
regionSouth     -2932.20     194.15 -15.103  < 2e-16 ***
regionWest            NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5531 on 5082 degrees of freedom
Multiple R-squared:  0.2351,    Adjusted R-squared:  0.234
F-statistic: 223.1 on 7 and 5082 DF,  p-value: < 2.2e-16
```

Figure 16. Model Summary from DiD Analysis

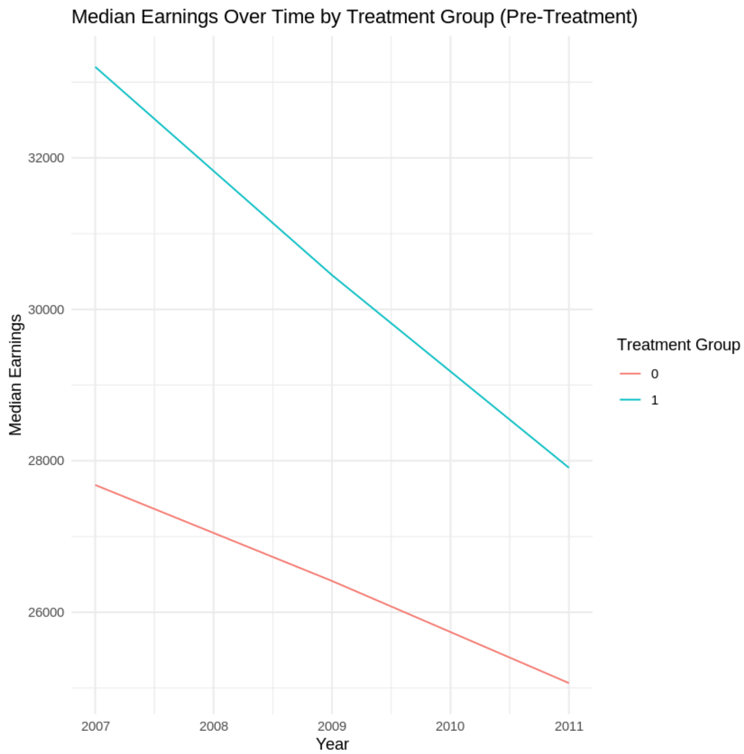

Figure 17. Graph to check Parallel Trends Assumption

```
[1] "ATT (DiD): 505.783582089555"
[1] "GATE estimates by region:"
work_ratio_binary:regionNortheast   work_ratio_binary:regionMidwest
                         1328.917                         -2496.624
    work_ratio_binary:regionSouth
                        -2924.168
```

Figure 18. ATT and GATE Results from Educational Quality Results

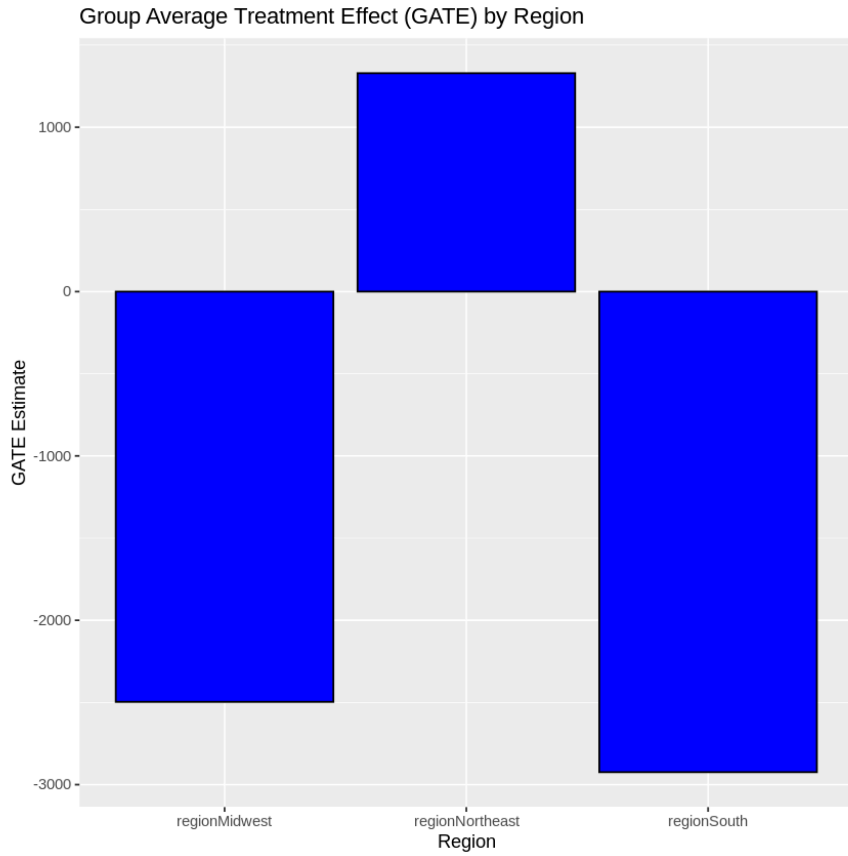

Figure 19. GATE Visualizations

Appendix C: Educational Level Analysis Results

Balance Measures

	Type	Diff.Adj
distance	Distance	-0.0000
work_ratio	Contin.	-0.0484
regionNortheast	Binary	-0.0346
regionMidwest	Binary	-0.0591
regionSouth	Binary	0.0949
regionWest	Binary	-0.0012
year_2007	Binary	-0.0290
year_2009	Binary	-0.0269
year_2011	Binary	-0.0312
year_2012	Binary	0.0232
year_2013	Binary	0.0226
year_2014	Binary	0.0413

Figure 20. Summary of Balance Measure of PSM for the the Education Level

Sample sizes

	Control	Treated
All	19467.	10978
Matched (ESS)	920.82	10978
Matched (Unweighted)	3128.	10978
Unmatched	16334.	0
Discarded	5.	0

Figure 21. Summary of Sample Sizes from PSM for Education Level Analysis

```
Call:
lm(formula = as.formula(paste(outcome_2, "~", treatment_2)),
    data = matched_data_logistic_2)

Residuals:
   Min      1Q Median     3Q    Max
-31515  -6715  -1215   4985 128585

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                     33845.5      192.7  175.68   <2e-16 ***
pred_degree_awarded_ipeds_binary 9470.0      218.4   43.36   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 22. Model Summary from PSM using Education Level

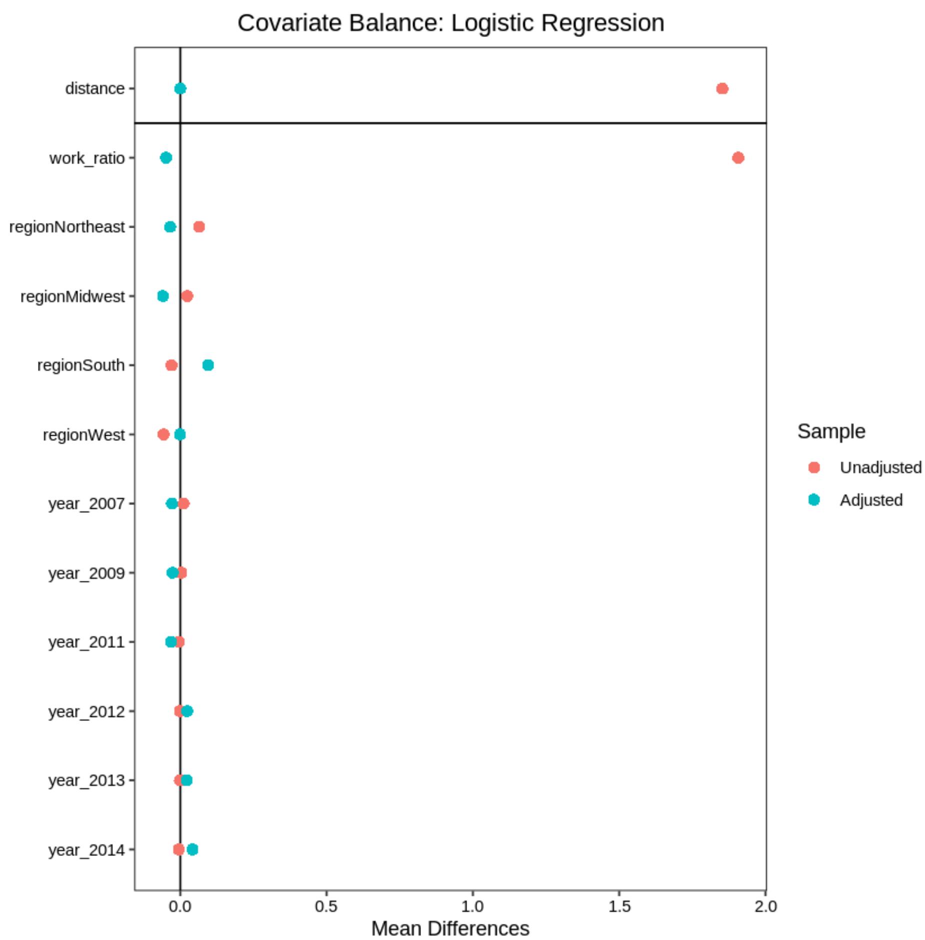

Figure 23. Love Plot from PSM using Education Level

```
       V1
 Min.   :0.00000
 1st Qu.:0.03277
 Median :0.20895
 Mean   :0.35511
 3rd Qu.:0.69758
 Max.   :0.99741
Warning message:
"Using one column matrices in `filter()` was deprecated in dplyr 1.1.0.
i Please use one dimensional logical vectors instead."
       V1
 Min.   :  0.4245
 1st Qu.:  0.8189
 Median :  2.2617
 Mean   :  8.9603
 3rd Qu.:  7.2072
 Max.   :957.0373
       V1
 Min.   :-18407592
 1st Qu.:    -7902
 Median :     -740
 Mean   :     5405
 3rd Qu.:     9188
 Max.   : 17292474
[1] "ATE (AIPW): 5405.08475055529"
```

Figure 24. IPW Results from Education Level Results

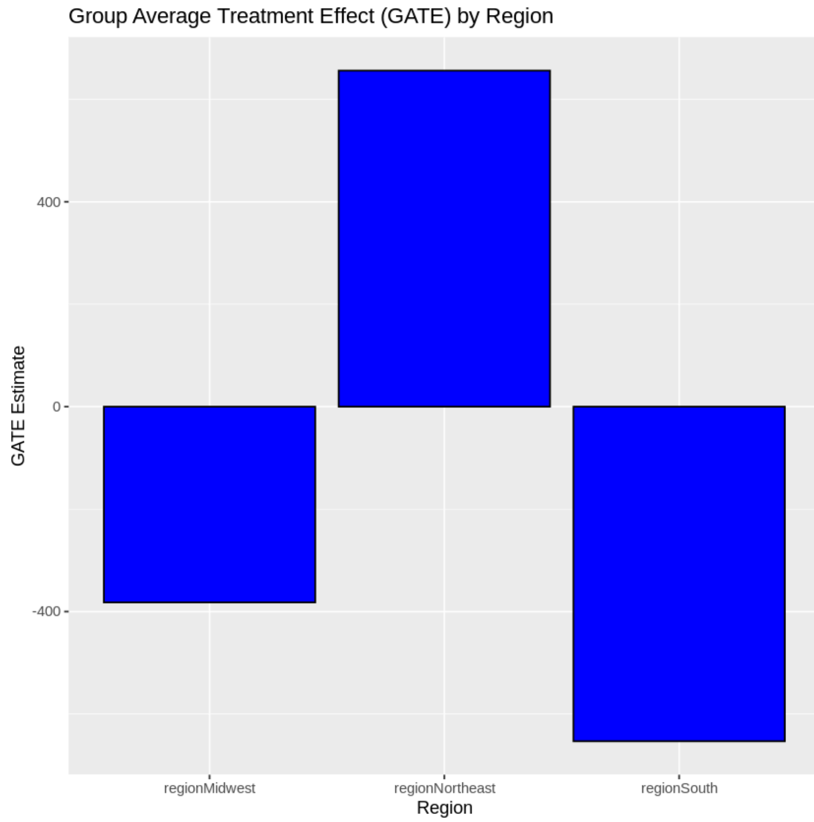

Figure 25. Graphical Visualization for GATE for Educational Level Results