

# **Analyzing the Monetary Effect of Going to College on Annual Earnings**

*Question:* Is going to college a worthwhile investment?

## **Introduction**

With some in the public beginning to question the value gained from attaining higher education, this report was written with a clear objective in mind – to use past data to analyze whether such claims were substantiated or merely beliefs rippling throughout popular society. To answer the question of whether going to college is a worthwhile investment or not, we used data gathered from the National Longitudinal Survey of Youth, conducted by the U.S. Department of Labor’s Bureau of Labor Statistics. The survey randomly selected and tracked roughly 6,000 men and women born between the years 1957 and 1964 and captured detailed records of each participant’s life and upbringing. However, for the purposes of this analysis, we chose a little under 4,000 individual observations to work with. Our hope with this report is that the analyses contained within it help explore the earnings gap between college graduates and non-graduates, explaining the impact various variables have on people’s earnings and allowing readers to understand the isolated monetary effect going to college has on their future earnings.

## **Descriptive Statistics on Earnings for College Graduates and Non-Graduates**

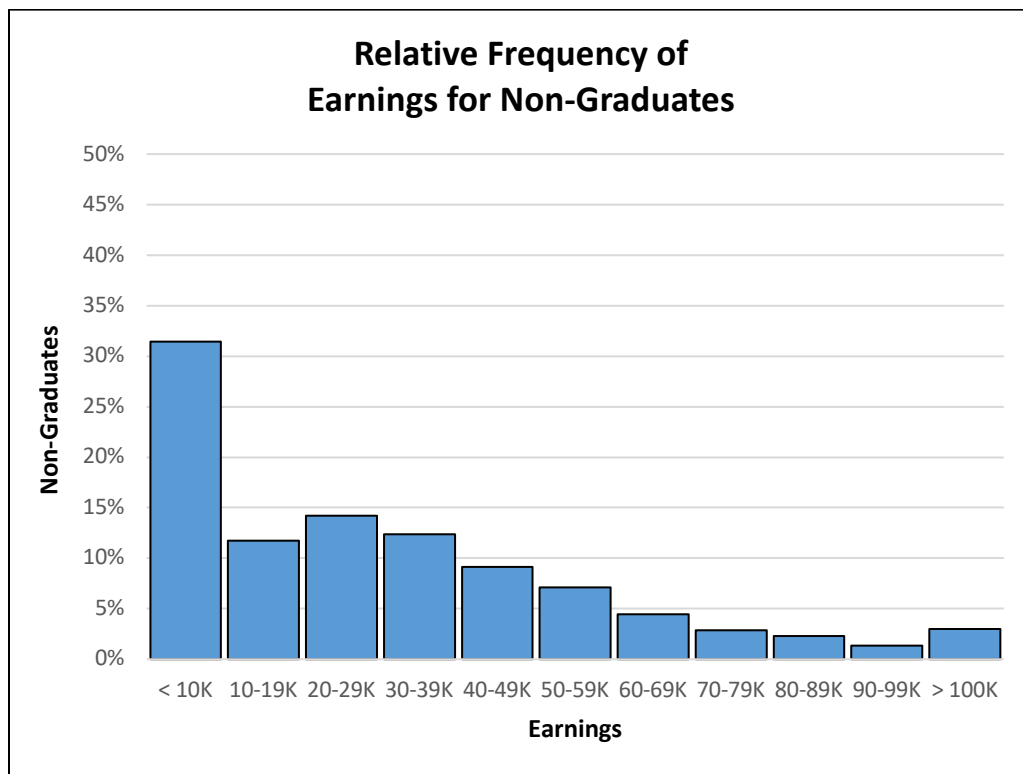
Before we begin analyzing the sources of the earnings gap between college graduates and non-graduates, it is helpful to understand what the earnings gap is without taking other variables into consideration. As shown in **Figure 1**, on average, non-college graduates earned \$31,881 per year, while graduates earned \$80,150 per year. This amounts to an on-average annual difference of \$48,269 in earnings. However, discussion of the mean is not complete without further analyzing the accompanying median value: the annual earnings of an individual located at the 50th percentile. Taken together, the mean and the median show relative skewness, or the existence and influence of outliers on the dataset. For non-graduates, the mean and median (\$25,000) are roughly close together, while for graduates, their difference is much larger. This indicates that outliers like those who made \$312,324 a year may have affected the average annual earnings of graduates more, although both graduate and non-graduate datasets appear to have outliers within them. Just taking medians into consideration, there still exists a substantial \$35,000 annual earnings gap between college graduates and non-graduates at the 50th-percentile level, appearing to show at a surface-level that graduates tend to earn more than non-graduates.

**Figure 1**

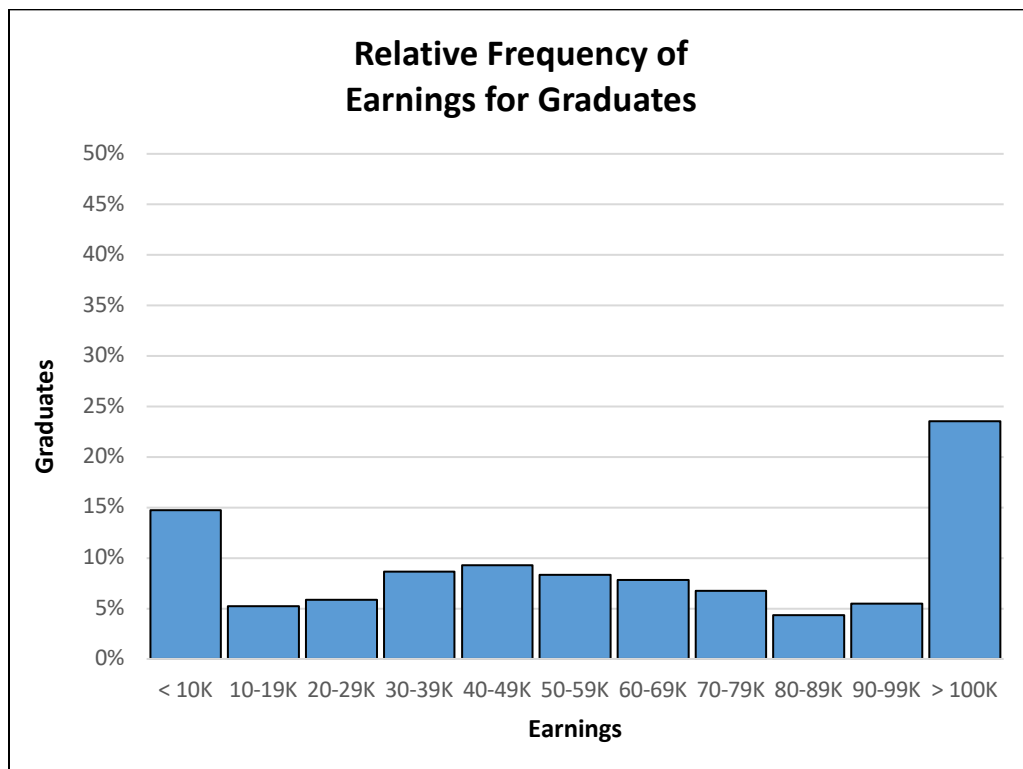
<b>Annual Earnings by College-Graduate Status</b>		
	<i>Non-graduate</i>	<i>Graduate</i>
Mean	\$31,881	\$80,150
Median	\$25,000	\$60,000
Standard Deviation	\$36,968	\$81,976
Minimum	\$0	\$0
Maximum	\$312,324	\$312,324
25th Percentile	\$2,400	\$30,000
75th Percentile	\$46,000	\$100,000
Count	2839	1078

Next, let us understand a visualization of the data just summarized. Below are **Figures 2a** and **2b**, the relative frequencies of earnings based on college graduation status. As shown in **Figure 2a**, the highest percentage (approximately 30%) of non-graduates earned between \$0 and \$10,000 annually. The next highest percentage is near 15%, which was in the \$20,000 to \$29,000 annual earnings range. Meanwhile, in **Figure 2b**, the highest percentage of college graduates (approximately 25%) earned more than \$100,000 annually, further indicating that the statistics shown earlier were not heavily influenced by a single outlier or small selection of outliers, but instead a considerable amount of those surveyed. There in fact appears to be a plurality of college graduates who earned much more than the mean. Next, for merely comparative purposes, approximately 15% of college graduates earned \$0 to \$10,000 annually. Taking into account the difference in the sizes of the two groups (non-college-graduates accounted for more than double the size of college graduates), we can safely conclude that more non-graduates earned less than \$10,000 a year than graduates did. Additionally, looking at the histograms themselves, we can trace the low average annual earnings among non-graduates to a high percentage of low or zero incomes reported, whereas for college graduates, we see a much righter skewed distribution of earnings. This appears to show that a higher percentage of college graduates break into the six-figure earnings range than non-graduates.

**Figure 2a**



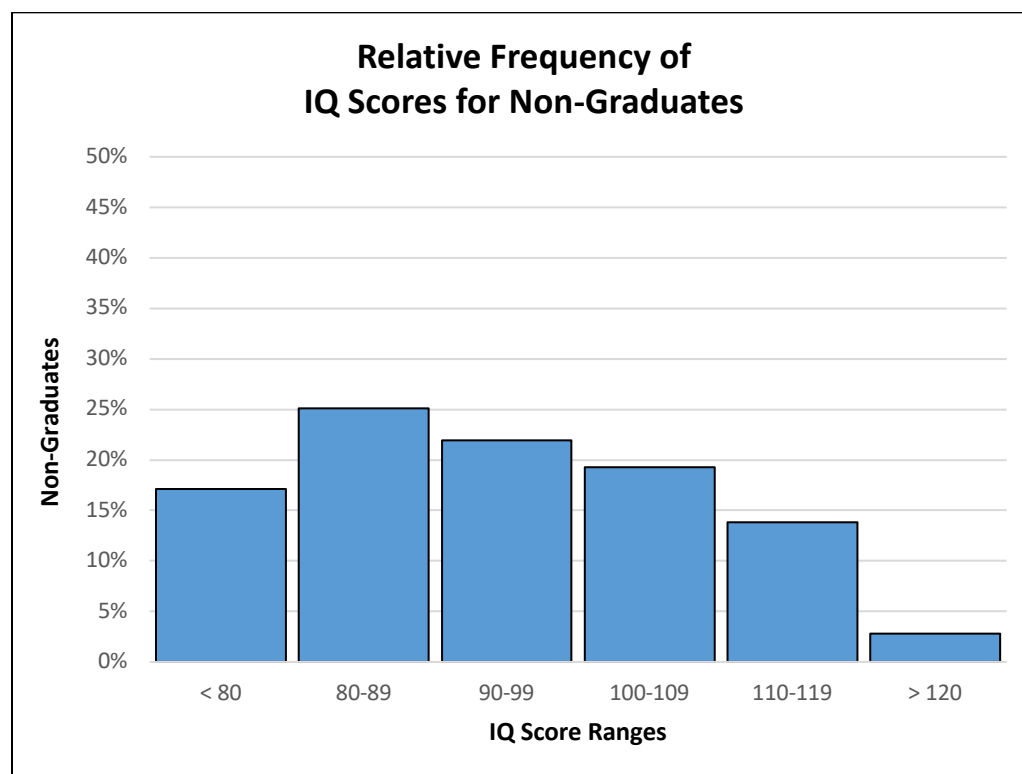
**Figure 2b**



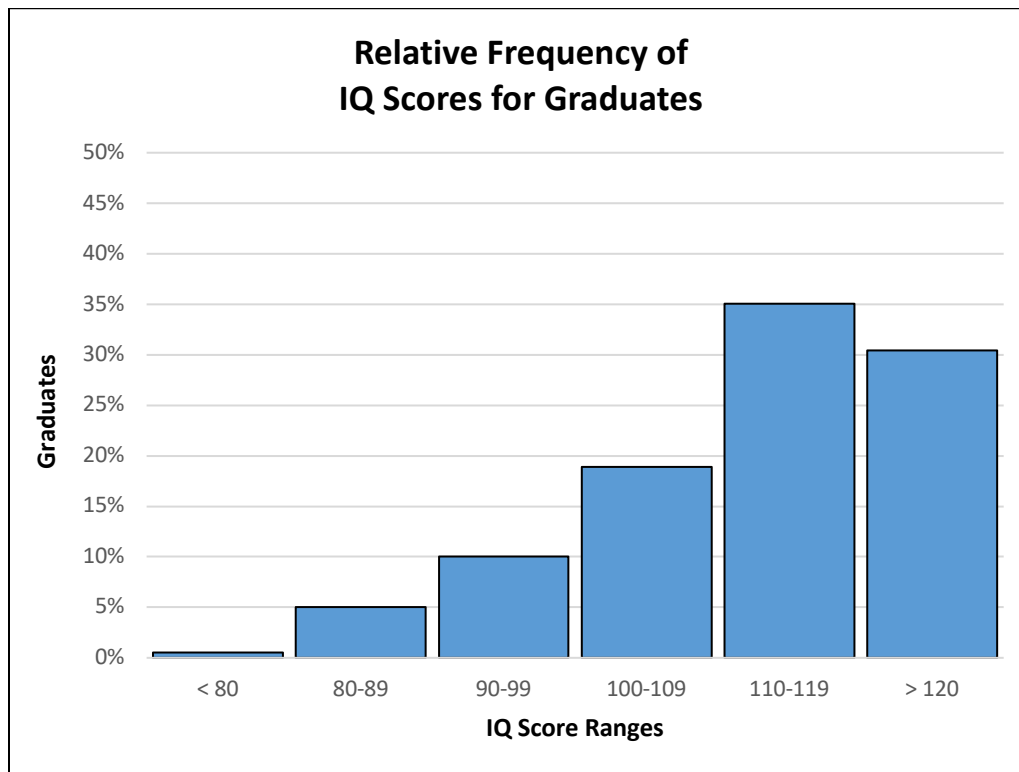
## Effects of IQ on College Graduate Status and Earnings

A hidden factor which may help to explain the sufficient earnings gap between college graduates and non-graduates is each group's relative frequency of IQ scores. As per **Figures 3a** and **3b** below, college graduates scored higher on IQ tests than non-graduates. 35% of graduates landed in the 110-119 IQ scoring range, while less than 15% of non-graduates achieved the same results. Similarly, while 25% of non-graduates scored between 80 and 89 on the IQ assessment, only 5% of graduates did the same. Going to college may not be the underlying factor causing this divide between the IQ scores of graduates and non-graduates. However, it may be that those who have higher IQs are more likely to go to college and graduate. IQ is a complex human cognitive trait that is influenced by both genetics and environmental factors surrounding a child growing up. By the time a person goes to college, their IQ is inherent. Even though a person may continue to learn, their cognitive abilities will most likely remain the same. Still, those with higher IQs may be able to accomplish more complex and brain-intensive tasks, which may factor into their earnings and the type of work they do.

**Figure 3a**



**Figure 3b**

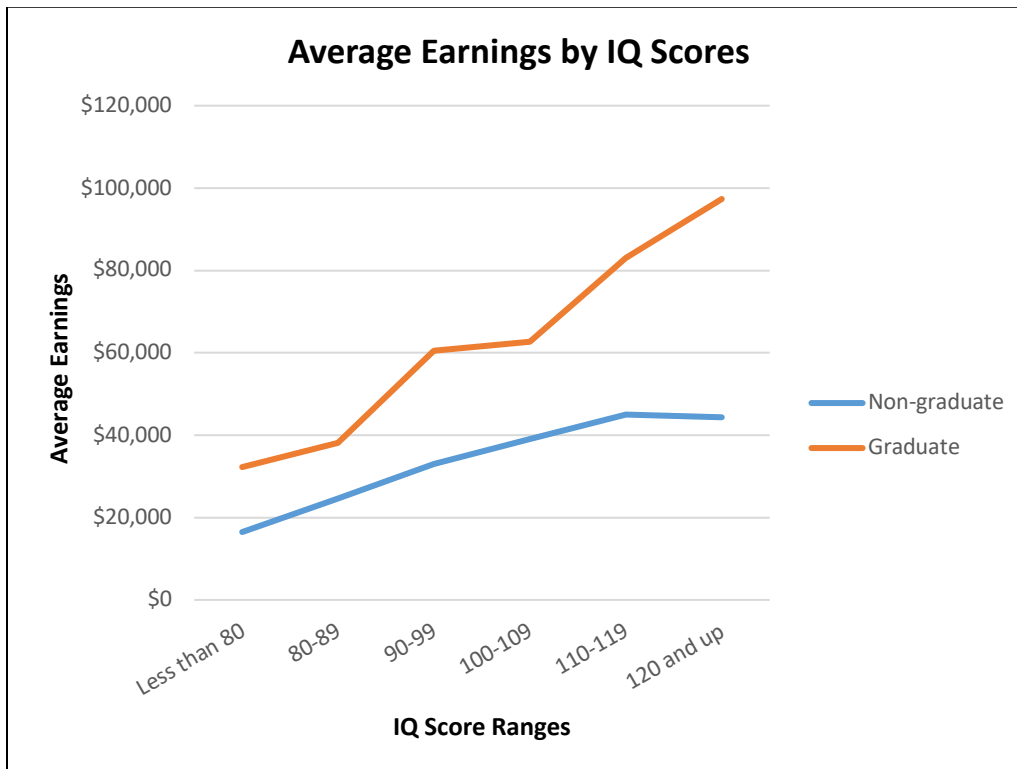


However, as **Figures 4 and 5** indicate, even for workers within the same IQ scoring range, those who were college graduates tended to earn much more than their non-graduate counterparts, appearing to show the additional value a college degree has. As shown in **Figure 5**, there is a positive relationship between a person's IQ and his/her's average annual earnings. However, the magnitude of the relationship differed between college graduates and non-graduates. Drawing on the data provided in **Figure 4**, college graduates who scored less than an 80 on the IQ test earned nearly \$16,000 more per year than their non-graduate counterparts. This trend continues for those scoring between 80 and 89, as they out-earned non-graduates by over \$13,000. On the other side of the spectrum, however, looking at those who scored 120 or higher on the IQ test, college graduates earned nearly \$53,000 more than equally intelligent non-graduates! This goes to show the importance of getting a college degree, and its supplementary educational value to natural intellect.

**Figure 4**

Average of earnings IQ Score Ranges	College Graduation Status	
	Non-graduate	Graduate
Less than 80	\$16,475	\$32,250
80-89	\$24,675	\$38,182
90-99	\$32,989	\$60,519
100-109	\$39,080	\$62,686
110-119	\$45,000	\$83,018
120 and up	\$44,391	\$97,340

**Figure 5**



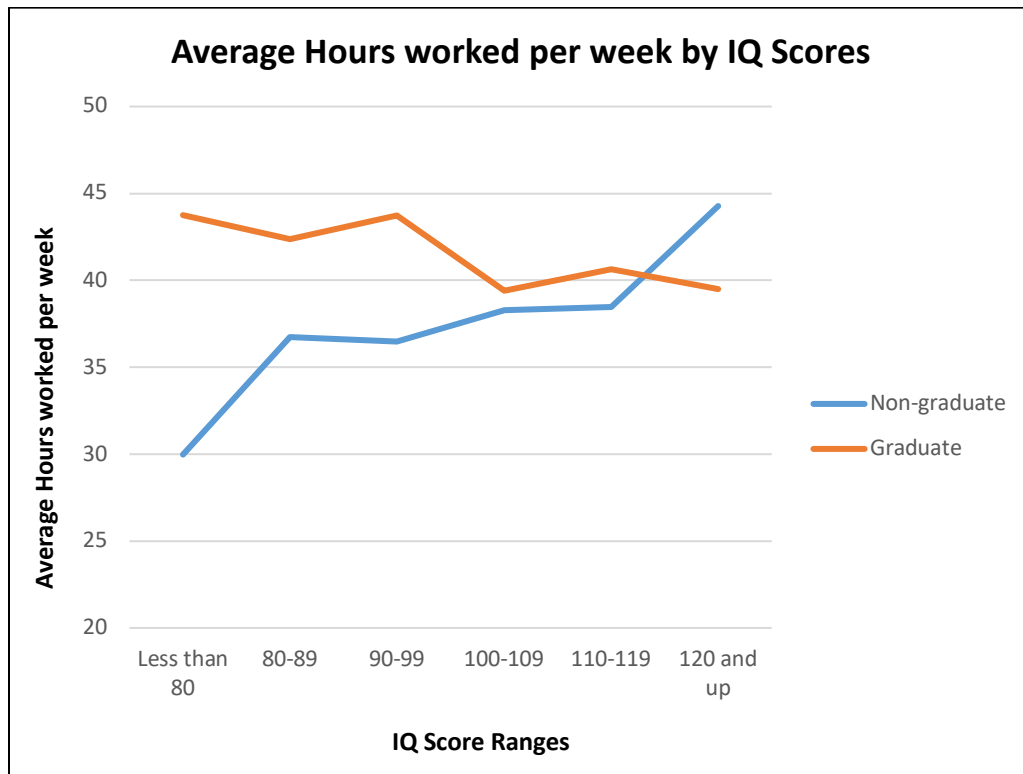
## **Effects of Weekly Hours Worked on Earnings**

Graduation status and IQ scores may also have an effect on the average number of hours an individual works per week, a key driver of overall annual earnings. **Figure 6** below tracks these variables in relation to hours worked per week. As shown, college graduates tended to work more hours per week than their non-graduate counterparts regardless of IQ scores. This provides an insight into one of the sources of the earnings gap. College graduates may earn higher hourly wages than non-graduates, so coupled with their higher average hours worked per week, college graduates may inherently earn more than non-graduates.

An individual may benefit from being a college graduate in more ways than just having higher annual earnings. Taking IQ scores into consideration, the higher IQ a college graduate had, the fewer hours he/she worked per week, perhaps allowing for more flexibility and relaxation in their schedule. However, in the case of non-graduates, the inverse appeared to be true. As per **Figure 6**, non-graduates whose IQ scores landed in the 120 and above range worked more hours per week than graduates, capping off a trend of rising weekly work hours with rising IQ scores. However, this might also be the effect of having a few outliers in the dataset. For example, a high-IQ student who dropped out of college to pursue a startup idea on a full-time basis would be included in this category, even though he/she is not representative of the entire population of non-graduates with high IQs. As previously suggested in the Descriptive Statistics section above, the data used for this analysis contains outliers which may affect the stated results, thus it is not definitive that non-college-graduates with high IQs work more hours than their college-graduate peers.



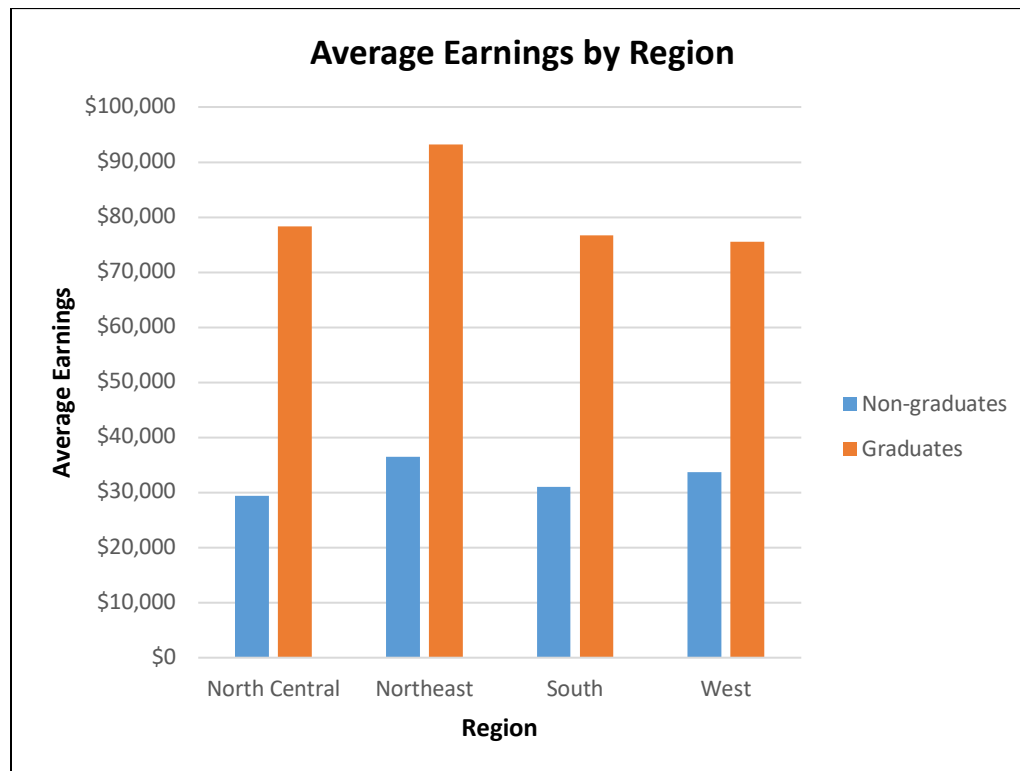
**Figure 6**



## Effects of Region on Earnings

When deciding on whether college is a worthwhile investment or not, another factor to consider for prospective college students is the earnings gap by region in the United States. As shown in **Figure 7**, the earnings gap between college graduates and non-graduates appears to be the greatest in the Northeast, with annual earnings differing by over \$56,000. However, in the South and West, the gap was much smaller at nearly \$46,000 and \$42,000, respectively. This chart of college graduate and non-graduate earnings also appears to show another phenomenon. While the earnings gap was the highest in the Northeast, graduates and non-graduates living in the region earned the most compared to their fellow graduate status peers living in other regions of the country. The reason for this may be because employers in the region have adjusted workers' salaries to keep up with the higher costs of living associated with the Northeast, regardless of whether workers are college educated or not.

**Figure 7**



## **Isolating the Monetary Effect of Being a College Graduate**

From our above analysis of the dataset and its characteristics, we have explored potential factors which we believe help explain the difference in average annual earnings between college graduates and non-graduates. Now, for the second part of this report, we will dive deeper into these factors to quantify the impact each variable has on the earnings gap and attempt to isolate the true effect of college graduation on a person's annual earnings.

### **Simple Regression Model**

First, we ran a simple regression (**Regression A in Figure 8**) to determine the effect being a college graduate has on a person's average earnings. As we expected (and actually proven earlier in the Descriptive Statistics section of this report), the average difference in earnings between college graduates and non-graduates came out to be roughly \$48,269. However, the Adjusted R-Squared statistic for this regression was around 0.14, suggesting that this regression only explained approximately 14% of the variation in annual earnings, clearly indicating that other variables besides college graduation status affect a person's annual earnings.

### **First Multiple Regression Model**

Next, we included new variable tests into the regression model (**Regression B in Figure 8**) in an attempt to account for more variation in annual earnings. First, we included a test for gender, as repeated research has shown that males typically earn more than females – a phenomenon known as the gender pay gap. We also included a test for determining the on-average effect of a person living in a rural area on earnings. Adding in these variables into the simple regression introduced in **Regression A** decreased the on-average effect of a person being a college graduate slightly, showing that the College\_Graduate coefficient in the simple regression may have been biased because it incorporated the effects of other variables that were omitted from the regression. However, with these new variables now factored into the regression model, some of the omitted variable bias present was removed. Despite the addition of the Male and Rural variables into the regression model, **Regression B** still explained approximately 20% of the variation in earnings, suggesting further variables needed to be added into the multiple regression model to find the true isolated effect of college graduation on earnings.

### **Second Multiple Regression Model**

Following our addition of the Male and Rural variables, for our next regression model (**Regression C in Figure 8**), we also included consideration for which IQ category individuals tested into, as we previously suggested that there may be an association between a person's cognitive ability and average annual earnings. The reason we categorized IQ scores into ranges of 10 is that we recognized that minute differences in individual IQ scores may not indicate significant differences in cognitive ability, whereas nearly double-digit differences may provide a better indication of any differences in earnings. With the inclusion of IQ categories, the

College\_Graduate coefficient reduced significantly, confirming that there was omitted variable bias in our earlier regression models. From the coefficients of this regression, we can interpret that, if a person is a college graduate, he/she will earn \$33,757.50 more on average compared to a non-graduate, holding all other variables such as gender, IQ category, and rural residence constant. Given that the coefficient is statistically significant (with a t-statistic larger than 2 in magnitude), we are 95% confident that a person's college graduation status has an effect other than \$0 on their annual earnings, and that this relationship was not a chance occurrence. Similarly, if a person is male, they earn \$27,937.41 more on average than a female would, holding all other variables constant. Since this too is statistically significant, we can be 95% confident that a person's gender has an effect other than \$0 on their annual earnings. On the other hand, our regression model states that if a person lives in a rural area, they will earn \$2,357.59 less on average, holding all other variables constant. However, this value is not statistically significant because \$0 falls within two standard deviations of the coefficient. We cannot be 95% confident that a person's rural area of residence has an effect other than \$0 on their annual earnings, thus suggesting this relationship may be a chance occurrence. This is verified by the rural variable's t-statistic value being less than two. With this information, we now recognize that rural residency status may not be a significant variable in explaining the earnings gap between college graduates and non-graduates, and we will not consider including it in future regression models to explain the college graduation earnings gap.

Special attention needs to be placed in analyzing **Regression C's** IQ category coefficients. If a person's IQ measures below an 80, on average, they are expected to earn \$20,924.80 less than those with IQs in the average range of 100 to 109, holding all other variables constant. Similarly, if a person's IQ is within the 80 to 89 range, they are expected to earn \$14,098.31 less than a person with an IQ in the range 100 to 109, holding all else constant. On the other side of the spectrum, if a person's IQ is in the range 110 to 119 (above average), they are expected to earn \$9,059.94 more on average than a person whose IQ is in the 100 to 109 range, holding all else constant. This difference more than doubles when discussing the earnings of individuals whose IQs are above 120. On average, people with an IQ greater than 120 earn \$20,478.31 more than those with IQs in the 100 to 109 range, holding all other variables constant. Since each of these coefficients is statistically significant, we can say with 95% confidence that having an IQ at least 10 points away from the average (100) has an effect other than \$0 on a person's annual earnings. The same cannot be said for people with an IQ score in the range of 90 to 99, because on average, they earn \$2,729.89 less than those with an IQ score between 100 and 109. Since \$0 falls within two standard deviations of this value and the coefficient's t-statistic is less than two in magnitude, we cannot say with 95% confidence that having an IQ score in the range of 90 to 99 has an effect other than \$0 on a person's annual earnings. This seems to indicate that those with higher-than-average IQ scores tend to earn more on average, while those with lower-than-average IQ scores tend to earn less on average, holding all other variables constant.

### **Omitted Variable Bias in the College Graduate coefficient**

Analysis of **Regression C**'s IQ category coefficients also confirms what was stated earlier in this report on the relationship between a person's intellect and annual earnings. While a causal relationship cannot be established between higher IQ scores and higher earnings, it can be suggested that those with higher IQ scores may work in more mentally intensive industries which may also tend to pay workers more. This sheds light on the relationship between cognitive ability, college attendance, and earnings. It may be that those with better cognitive ability (higher IQ scores) are more likely than their less-cognitively-gifted peers to attend college, and similarly, it may be that those with higher cognitive ability are more likely than others to attain higher earnings. This explains why the College\_Graduate coefficient dropped significantly after the inclusion of IQ categories into the regression. Before, in **Regressions A** and **B**, it appears that the College\_Graduate coefficient not only included the direct effect of graduating from college on a person's average annual earnings, but also incorporated the indirect effects of a person's IQ category on their average earnings, since the background relationship suggests that cognitive abilities affect a person's earnings. Once the IQ categories were added into the regression, the omitted variable bias present in earlier regressions became clear, showing that we were overestimating the true effect of being a college graduate on annual earnings.

**Figure 8****Regressions of College Graduate Status on Earnings\***

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
College_Graduate	48268.69** (1906.30)	48135.96** (1846.90)	33757.50** (2151.98)	22304.16** (2917.19)
Male		28715.60** (1642.52)	27937.41** (1615.43)	25010.09** (1581.94)
Rural		-2256.96 (1753.48)	-2357.59 (1717.42)	
IQ_Less_80			-20924.80** (3135.19)	-18537.93** (3121.29)
IQ_80_89			-14098.31** (2601.84)	-14798.57** (2563.44)
IQ_90_99			-2729.89 (2615.14)	-3146.65 (2550.42)
IQ_110_119			9059.94** (2620.58)	4277.63 (3091.11)
IQ_Greater_120			20478.31** (3121.86)	2319.06 (5119.49)

\*Continued on next page

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
ColGrad_IQ_110				15077.81** (4610.25)
ColGrad_IQ_120				30144.18** (6092.73)
Hours				702.54** (49.90)
Hours^2				-3.30** (0.34)
North_Central				-9723.96** (2409.88)
South				-4633.31** (2311.75)
West				-6036.93** (2643.75)
Constant (intercept)	31880.90 (1000.05)	19093.89 (1395.14)	24611.32 (2140.37)	13954.82 (2960.31)
Observations	3917	3917	3917	3917
SEE	53285.15	51318.62	50249.89	48728.16
Adjusted R-Squared	0.1405	0.2028	0.2356	0.2812

Standard errors in parentheses; \*\*|t-stat| > 2;

Omitted IQ Range 100-109, Northeast region of United States

## **Final Multiple Regression Model**

After performing regression analysis on a person's gender, college graduate status, rural residence, and cognitive ability (measured through IQ category), we believed there were more omitted variables which could help explain the college graduation earnings gap. To isolate the effect college graduation has on a person's earnings, we first made a correlation matrix to visualize the individual associations found between variables in our survey data. As shown in **Figure 9**, the following variables had what we interpreted as relatively strong correlations with earnings: if a person's mother went to college, if a person is a college graduate, if a person is married, how many weekly hours a person works, a person's gender, and a person's IQ. From there, we deduced which variables we best believed were relevant to explaining a person's earnings and selected to run a series of regression tests on combinations of these variables, each time attempting to build a regression model which we hoped would explain the highest percentage of variations (the highest Adjusted R-Squared statistic) in annual earnings. The regression we determined that best isolates the effect of college graduation on a person's annual average earnings is **Regression D** in the above **Figure 8**. From our analysis of the correlation matrix, previous regressions, and our own intuition, we believe the following factors best estimate the effect of college graduation on a person's earnings: college graduate status, gender, cognitive ability (measured through IQ categories), hours worked, and region of residence.

This regression model states with 95% confidence that if a person is a college graduate, they earn at least \$22,304.16 more on average than their non-college-graduate peers, holding all else constant. Furthermore, the regression suggests that even though people may have higher than average cognitive ability which may aid them in attaining higher earnings, those who also go to college tend to earn between \$15,077.81 and \$30,144.18 more on average than their cognitively equal peers, holding all else constant. This is shown through the coefficients on the interaction variables ColGrad\_IQ\_110 and ColGrad\_IQ\_120, confirming what was observed in **Figure 5**.

We also suspected that the average number of weekly hours a person works and their annual earnings do not form a linear relationship. Therefore, we included an Hours-Squared variable into our regression, and our hypothesis was soon confirmed when the coefficient on the squared-variable term was significant. Because of this, we can be confident that the relationship between the hours worked and annual earnings is nonlinear and in an inverted U-shape. This indicates that there are decreasing returns for hours worked – that after a certain number of hours, an additional hour worked may not increase earnings in the same way.

While our **Regression D** attempts to include all variables from the survey which we believe affects a person's annual earnings, we also recognize that other variables not included in the regression also play a role in determining earnings. For example, we presume that age and experience are both valued in the workplace, but the nature of our surveyed dataset provides a limited age range for us to run a regression on. Since all participants in the survey are known to be born between 1957 and 1964, variations in age may not seem to explain the earnings gap present in this dataset.



## **Sources of Bias in our Data**

Despite our best efforts to explain more of the variation in earnings through repeated regression modelling, none of our regression tests could explain more than 30% of the variation in a person's average annual earnings. As seen through its Adjusted R-Squared value, **Regression D** only explains approximately 28% of all variations in earnings. We believe this may not be a flaw with our logic or additional omitted variables, but perhaps an outcome influenced by the presence of outliers in the dataset. For example, the survey included data on individuals who worked zero hours a week but reported more than zero dollars in earnings. Similarly, there were individuals surveyed who reported working several hours a week with zero annual earnings or working an unusual number of hours beyond the number of hours available in a week. If we were to redo this analysis in the future, we would consider omitting these observations to boost the accuracy of our predictions.

This survey data also lends itself to the potential for self-selection bias, a phenomenon which results in a non-representative sample. Since the data in this survey is of individuals who were born in the late 1950s and went to college in the late 1970s to early 1980s, we cannot accurately predict present earnings of college graduates based on changes in preferences which have occurred since then. The racial, ethnic, and gender breakdown of students who went to school in the 1970s is not representative of the breakdowns found in prospective college applicants today. For example, the percentage of women who went to college back then is presumably much less than the percentage of women in the population who are thinking about going to college now. Moreover, with the price of college becoming increasingly more expensive, prospective students are more selective in their decisions to attend college than in the 1970s, when college was comparatively much cheaper and more people could afford to go without weighing the considerable monetary burden such a decision would take on their finances. As a result, individuals observed in this survey may not be representative of current trends involved in weighing the monetary benefits of going to college, as shown through analyzing the effect college graduation has on their earnings.

Furthermore, the individuals surveyed for this data may tend to be mid-career, established professionals who might have acquired additional skills through their workplace experiences, thus boosting their earnings through knowledge not acquired from college. To get a better understanding of the true effect college graduation has on a person's earnings, for future analysis, we suggest considering data from recent entrants into the workforce to better represent the effect a college degree has on earnings.

**Figure 9****Correlation Matrix\***

	<i>earnings</i>	<i>mother_college</i>	<i>college_graduate</i>	<i>age</i>	<i>married</i>
earnings	1				
mother_went_to_college	0.18	1			
college_graduate	0.38	0.33	1		
age	-0.01	0.02	0.01	1	
married	0.11	0.04	0.11	0.02	1
hours	0.20	-0.01	0.06	-0.02	0.01
Male	0.25	-0.01	0.00	-0.03	-0.04
Rural	-0.06	-0.10	-0.11	0.01	0.08
IQ_Less_80	-0.17	-0.13	-0.20	0.01	-0.10
IQ_80_89	-0.17	-0.16	-0.24	0.00	-0.09
IQ_90_99	-0.07	-0.09	-0.15	0.00	-0.03
IQ_100_109	0.00	0.00	-0.02	0.01	0.05
IQ_110_119	0.15	0.14	0.20	0.00	0.09
IQ_Greater_120	0.27	0.25	0.43	-0.01	0.07
North Central	-0.03	-0.03	-0.01	-0.01	0.04
South	-0.03	-0.05	-0.04	-0.01	0.00
West	0.00	0.08	0.01	0.01	-0.03
Northeast	0.07	0.01	0.04	0.01	-0.01

\* See Appendix for full Correlation Matrix

## **Conclusion**

In conclusion, our statistical analysis suggests that college graduates make at least approximately \$22,304 more on average than non-graduates per year, holding all other variables constant, although this number is significantly higher for those with higher levels of cognitive ability. However, as Google is quick to point out, going to college is not the ultimate determinant of success for their concern. Instead, our analysis suggests that graduating from college comes with its monetary benefits, and prospective students must take these into consideration when making a decision. After all, the decision to go to college is a personal one, influenced by many factors which go beyond the scope of this report. However, to answer the question presented at the top, we believe that if a person intends to work, investing in a college education is a worthwhile venture, which repays its upfront cost in the years to come.

## Appendix

Figure 9

### Correlation Matrix

	earnings	mother_college	college_graduate	age	married	hours	Male	Rural	IQ_Less_80	IQ_80_89	IQ_90_99	IQ_100_109	IQ_110_119	IQ_Greater_120	North Central	South	West	Northeast
earnings	1																	
mother_went_to_college	0.18	1																
college_graduate	0.38	0.33	1															
age	-0.01	0.02	0.01	1														
married	0.11	0.04	0.11	0.02	1													
hours	0.20	-0.01	0.06	-0.02	0.01	1												
Male	0.25	-0.01	0.00	-0.03	-0.04	0.13	1											
Rural	-0.06	-0.10	-0.11	0.01	0.08	-0.01	-0.01	1										
IQ_Less_80	-0.17	-0.13	-0.20	0.01	-0.10	-0.09	0.01	0.01	1									
IQ_80_89	-0.17	-0.16	-0.24	0.00	-0.09	-0.01	0.02	0.02	-0.17	1								
IQ_90_99	-0.07	-0.09	-0.15	0.00	-0.03	0.00	-0.07	0.03	-0.16	-0.24	1							
IQ_100_109	0.00	0.00	-0.02	0.01	0.05	0.02	-0.02	0.00	-0.17	-0.24	-0.24	1						
IQ_110_119	0.15	0.14	0.20	0.00	0.09	0.03	0.03	-0.02	-0.17	-0.24	-0.23	-0.24	1					
IQ_Greater_120	0.27	0.25	0.43	-0.01	0.07	0.04	0.05	-0.06	-0.13	-0.19	-0.18	-0.19	-0.18	1				
North Central	-0.03	-0.03	-0.01	-0.01	0.04	0.00	0.01	0.06	-0.07	-0.02	0.02	0.00	0.03	0.03	1			
South	-0.03	-0.05	-0.04	-0.01	0.00	0.00	-0.03	0.12	0.12	0.08	-0.02	-0.02	-0.07	-0.08	-0.48	1		
West	0.00	0.08	0.01	0.01	-0.03	0.01	0.01	-0.12	-0.03	-0.05	0.01	0.02	0.02	0.02	-0.29	-0.36	1	
Northeast	0.07	0.01	0.04	0.01	-0.01	-0.02	0.01	-0.10	-0.04	-0.03	0.00	0.00	0.03	0.05	-0.28	-0.34	-0.21	1