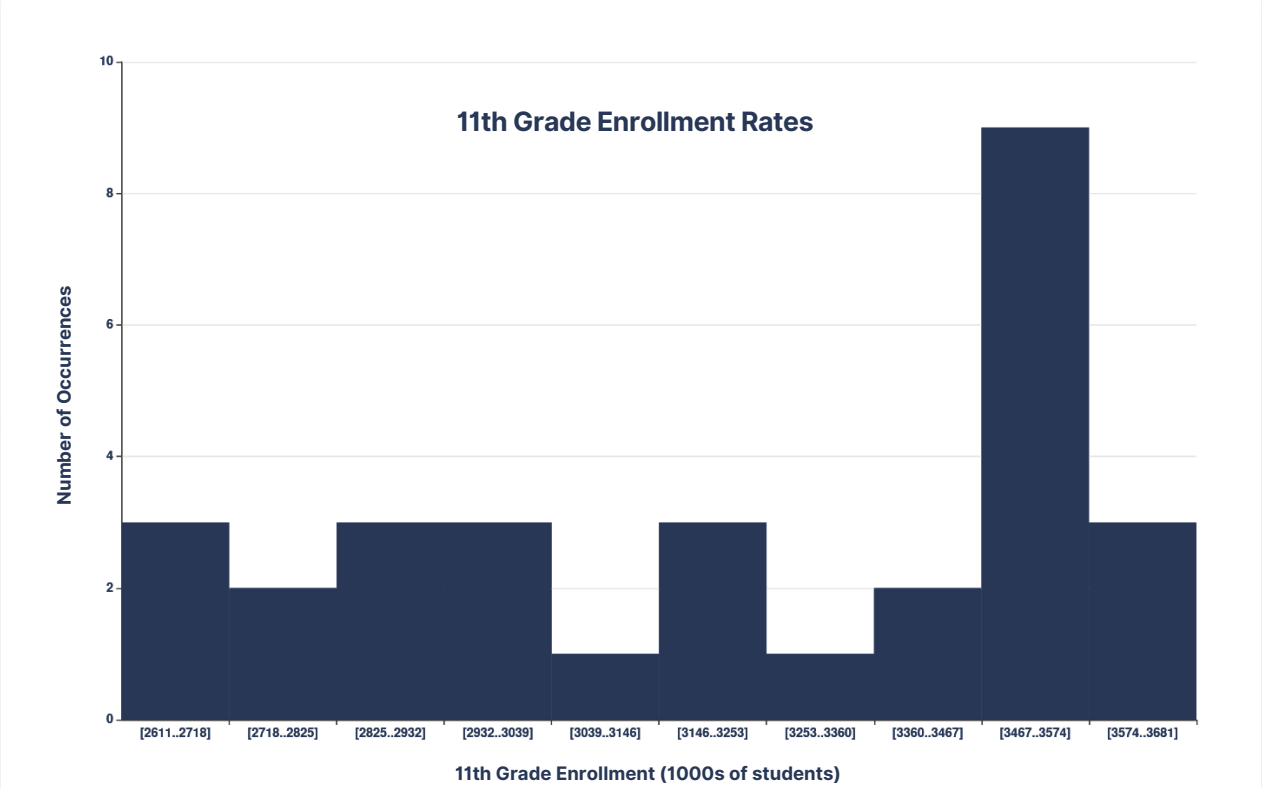
11th Grade Enrollment

Project by Nicholas Chiang and Maya Glazer



ONE VARIABLE ANALYSIS

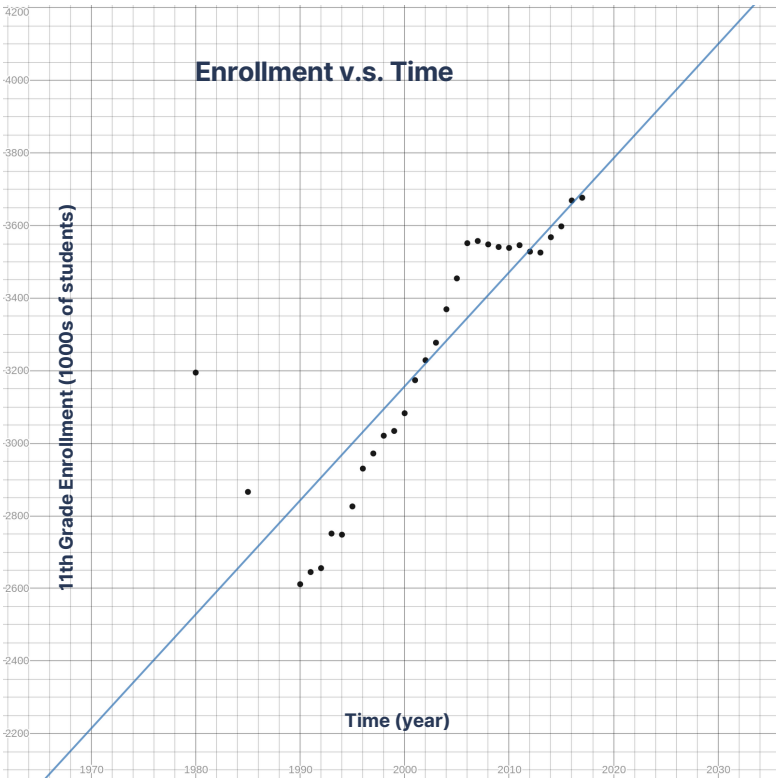
The median of 11th grade enrollment rates is 3,253,012 students (and the mean is 3,223,043 students). Since the mean (at 3,223,043 students) is less than the median (at 3,253,012 students), we conclude that enrollment rates for 11th grade are slightly skewed left; most of the data is towards the right while a few low data points skew the mean left. The interquartile range (IQR) for 11th grade enrollment is 603,918 students; there is a 603,918 student difference between the top and bottom of the middle 50% of our data. The standard deviation (SD) for 11th grade enrollment is 350,751 students; on average, data points are 350,751 students away from the mean.

3.2M
MEDIAN

603K
INTERQUARTILE RANGE

3.7M
MAXIMUM

Using the outlier formula, we discovered no outliers; there were no data points above *Q3 + 1.5 × IQR = 4,450,499* students (i.e. no high outliers) nor were there any data points below *Q1 - 1.5 × IQR = 2,034,826* students (i.e. no low outliers). Overall, the enrollment rates for 11th grade tend to be higher.



TWO VARIABLE ANALYSIS

By inspecting our scatterplot and visualizing an oval that includes all the data points in that scatterplot, we conclude that a moderately strong positive linear correlation between time and 11th grade enrollment rates exists (i.e. that oval is elongated and sloped upwards despite the two influential points we see towards the x-axis).

We modeled the relationship between 11th grade enrollment rates and time using a least-squares regression line (LSRL) with an equation of $\hat{y} = -59687.8 + 31.4224x$ where x is the year and $\hat{y}$ is the predicted enrollment rate for 11th grade in 1000s of students. Using this model, we predict that every year, 11th grade enrollment rates will increase by about 31,422 students.

After further analysis, however, we conclude that our least-squares linear regression line (LSRL) is not an appropriate model for this data set. While both the coefficient of determination (r^2) at 0.7369 indicates that 73.69% of the variation in 11th grade enrollment can be explained by time *and* the correlation coefficient (r) at 0.8584 reveals a moderately strong positive linear correlation between 11th grade enrollment and time, by inspecting the residual plot, we noticed a visible pattern, which demonstrates that a LSRL is not the best model for this relationship.

SUMMARY

We discovered that generally, as time went forward, 11th grade enrollment rates increased. We believe that this increase is primarily the result of an increasing population. Between 2006 and 2012, however, change in 11th grade enrollment rates remained stagnant and actually decreased slightly over time. We attribute this decline to the 1990s economic recession which led to a decrease in birth rates within the US: less children were born between 1989 and 1996 who would have been 11th grade students (16-17 years old) between 2006 and 2012.

REFERENCES

1. Digest of Education Statistics, 2019. (2017). National Center for Education Statistics; National Center for Education Statistics. https://nces.ed.gov/programs/digest/d19/tables/dt19_203.10.asp
2. Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Mann, F. B., Barmer, A., Dilig, R., Nachazel, T., Barnett, M., & Purcell, S. (2020). The Condition of Education 2020. National Center for Education Statistics. https://nces.ed.gov/pubs2020/2020144.pdf
3. Mather, M. (2012, July 18). The Decline in U.S. Fertility – Population Reference Bureau. Population Reference Bureau. https://www.prb.org/us-fertility/