

# Exploratory Data Analysis Project

Rachel Doering

9/18/2020

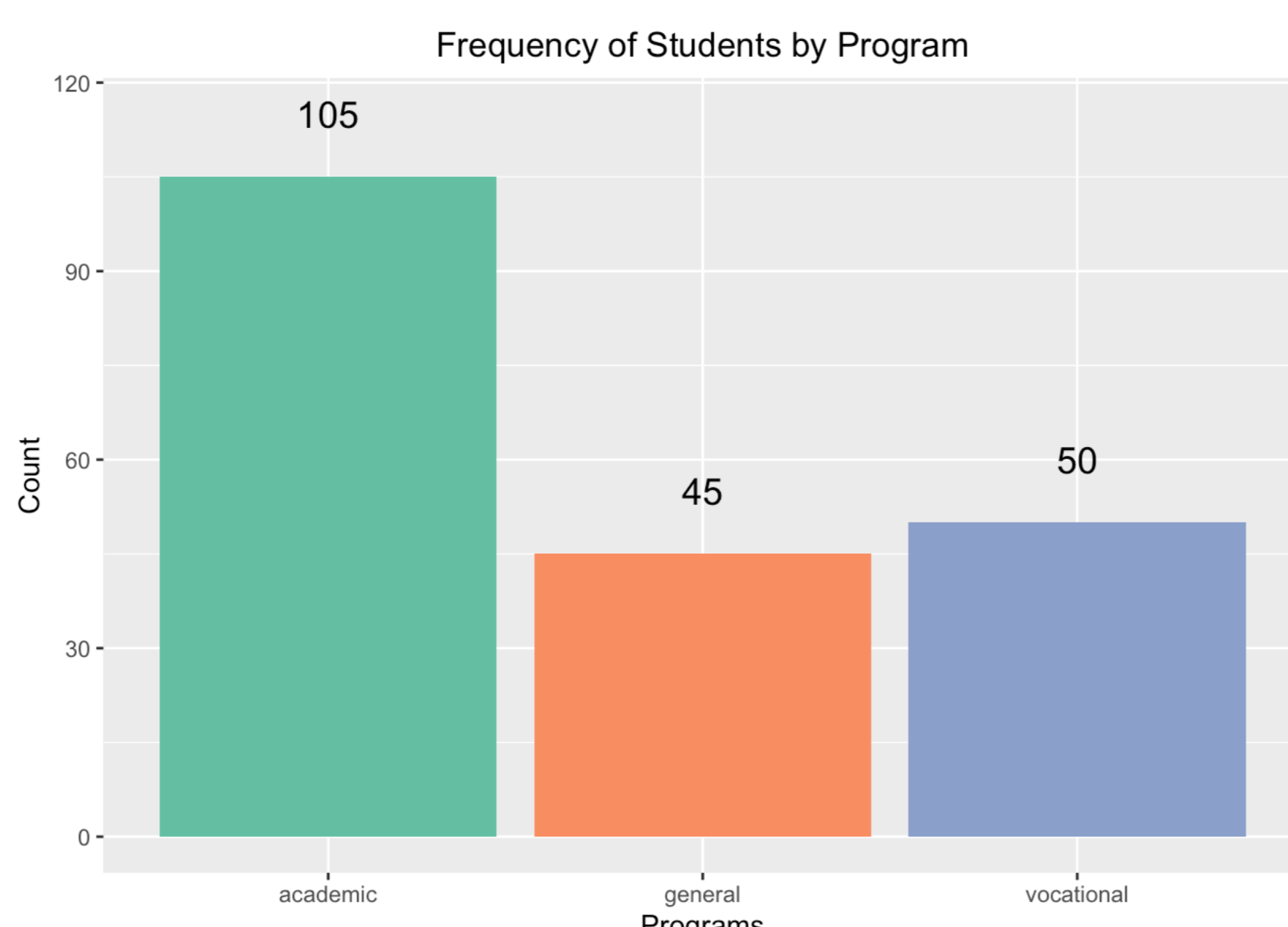
## Introduction - Highschool and Beyond

The High School and Beyond (HS&B) Longitudinal Study was the second study conducted as part of NCES' National Longitudinal Studies Program. This program was established to study the educational, vocational, and personal development of young people, beginning with their elementary or high school years and following them over time as they take on adult roles and responsibilities. In my personal analysis, I will be specifically looking at the different program types (prog) and how they take affect on math and science scores. There is definitely a stigma around vocational learning, or hands-on learning, being taught to younger groups so I am curious to see if they have the assumed disadvantage in STEM based curriculum.

```
highschool <-read.table("/Users/racheldoering/Desktop/MATH130/data/hsb2.txt", header=TRUE, sep="\t")
```

## Univariate Descriptions

```
ggplot(highschool, aes(x=prog, fill=prog)) + geom_bar(aes(y=..count..), show.legend = FALSE) + theme_gray() + ggtitle("Frequency of Students by Program") + xlab("Programs") + ylab("Count") + geom_text(aes(y=..count.. + 10, label=..count..), stat='count', size = 5) + scale_fill_brewer(palette="Set2") + theme(plot.title = element_text(hjust = 0.5))
```



The majority of this random sample of young people is enrolled in academic learning with a count of 105. There are 45 enrolled in general learning and 50 in vocational learning.

```
summary(highschool$math)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   \n##      33.00  45.00   52.00  52.65  59.00   75.00
```

The mean or average of math test scores overall is 52.65 with a minimum score of 33 and a max score of 75. The IQR is 14.

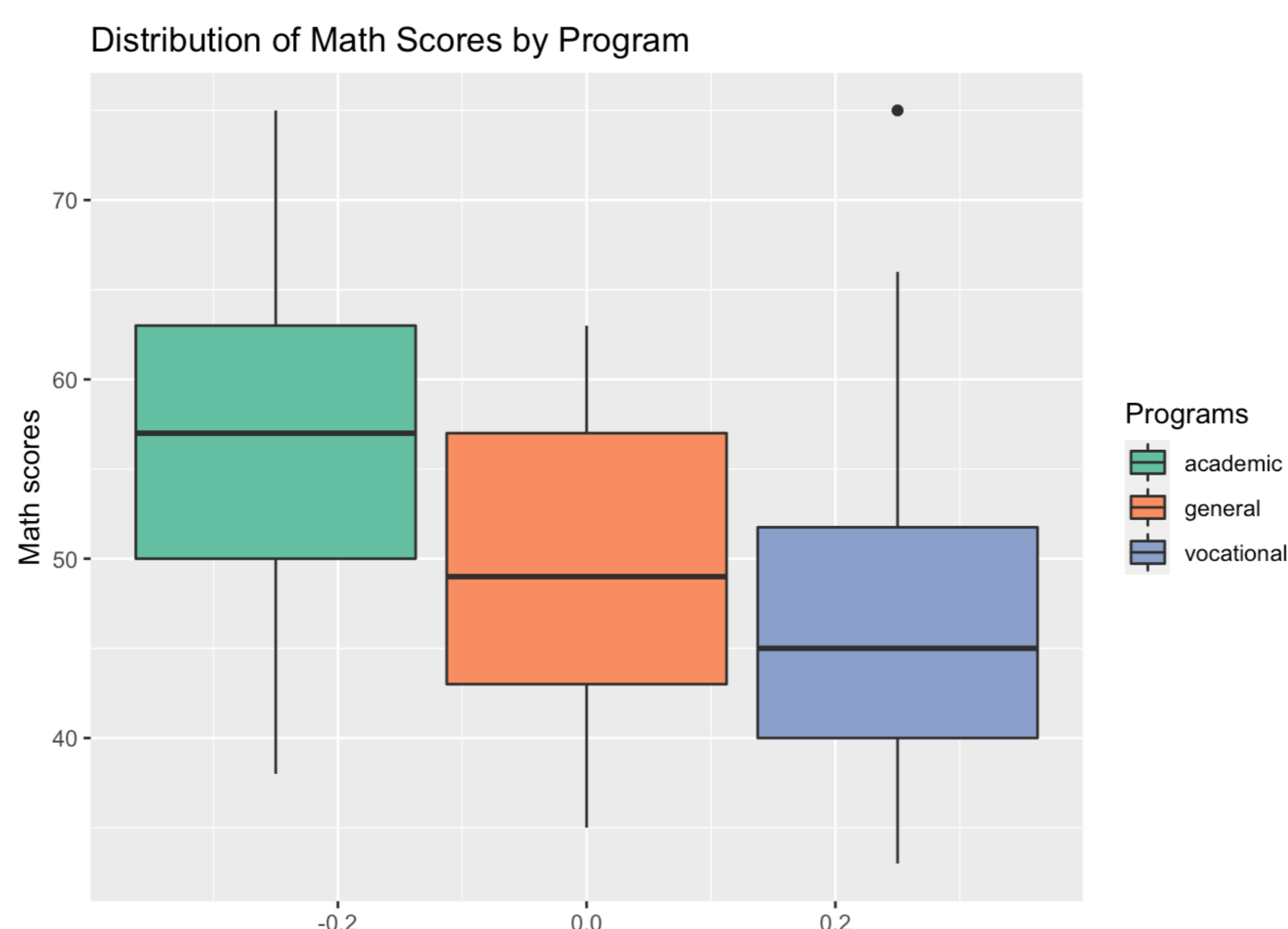
```
summary(highschool$science)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   \n##      26.00  44.00   53.00  51.85  58.00   74.00
```

The mean or average of science test scores overall is 51.85 with a minimum score of 26 and a max score of 74. The IQR is 14.

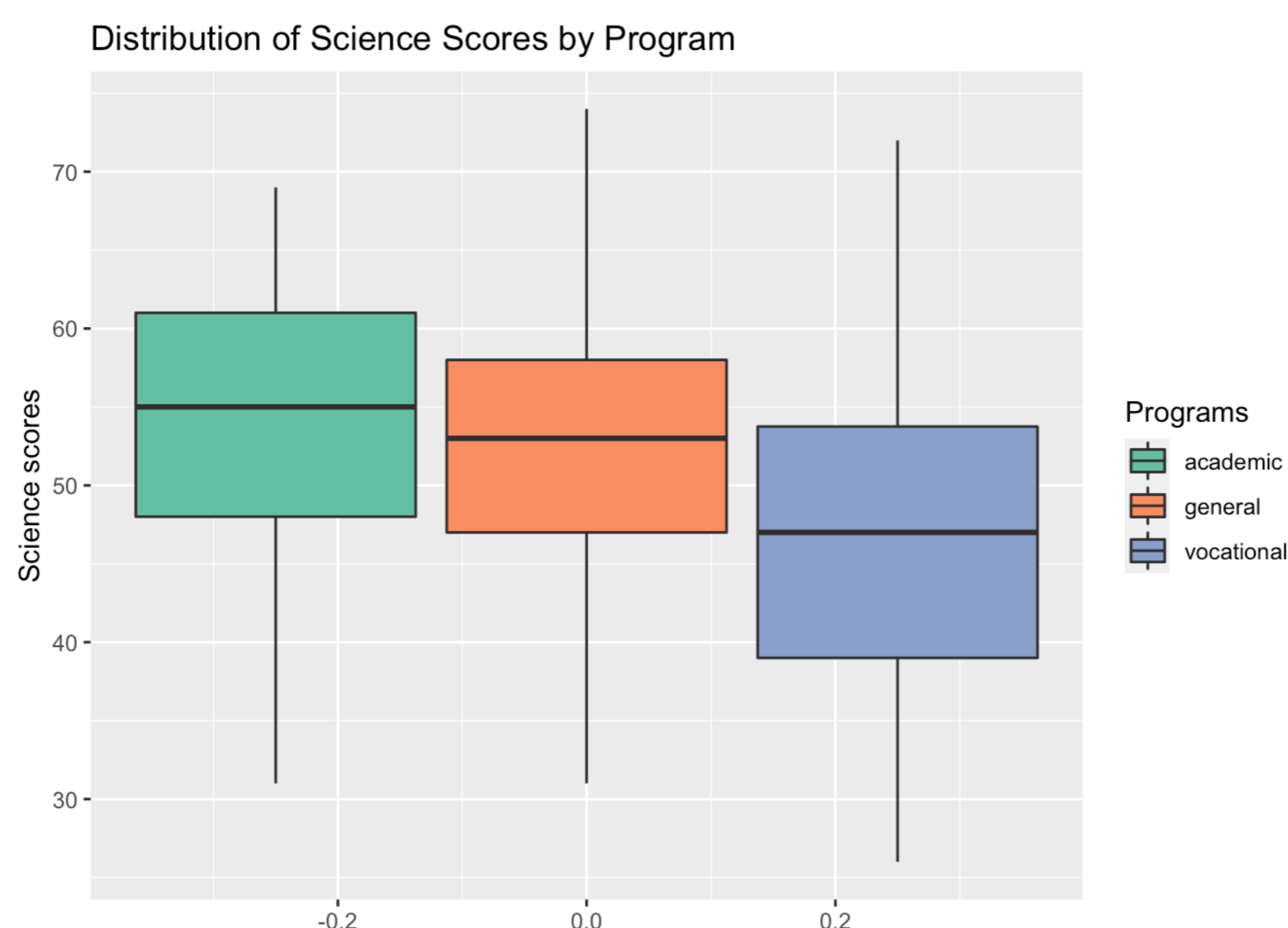
## Bivariate Comparison

```
ggplot(highschool, aes(x=math, fill=prog)) + geom_boxplot() + coord_flip() + scale_fill_brewer(palette="Set2", name="Programs") + ggtitle("Distribution of Math Scores by Program") + xlab("Math scores") + theme_gray()
```



This boxplot shows a higher median in math scores by those who were enrolled in academic learning with some overlap in the IQR. It also reveals an outlier of a much higher math score by someone enrolled in vocational learning appearing to be the same as the max score for academic learning scores.

```
ggplot(highschool, aes(x=science, fill=prog)) + geom_boxplot() + coord_flip() + scale_fill_brewer(palette="Set2", name="Programs") + ggtitle("Distribution of Science Scores by Program") + xlab("Science scores") + theme_gray()
```



This boxplot shows a slightly higher median in science scores for academic learning. It also shows a larger overlap in the IQR of the different programs. The max score for academic learning is lower than both general and vocational learning

```
table(highschool$prog, ifelse(highschool$math > 52.65, "above average", "below average")) %>% prop.table(margin=1)
```

```
##      \n##      above average below average \n## academic 0.6571429 0.3428571 \n## general 0.4444444 0.5555556 \n## vocational 0.2000000 0.8000000
```

This proportion table shows that 65.7% of those enrolled in an academic program scored above the overall average score of 52.65 in math. For the general program, 44% scored above average. In the vocational program, only 20% scored above average.

```
table(highschool$prog, ifelse(highschool$science > 51.85, "above average", "below average")) %>% prop.table(margin=1)
```

```
##      \n##      above average below average \n## academic 0.6380952 0.3619048 \n## general 0.5111111 0.4888889 \n## vocational 0.3200000 0.6800000
```

This proportion table shows that 63.8% of those enrolled in an academic program scored above the overall average score of 51.85 in science. For the general program, 51% scored above average. In the vocational program, 32% scored above average.

## Conclusion

Based on the bivariate comparison of program type with math and science scores, it appears that overall, those who were enrolled in academic learning scored better in math and science than those enrolled in general and vocational learning. This is most likely due to the fact that academic and general learning is knowledge based, focused on student assessment. Vocational is experience based, focused on students ability to perform tasks.