Introduction

As a group of data scientists, we are curious about how youths' behavior can affect their physical and educational outcomes. We wondered how specific behaviors related to technology use, exercise frequency, and drug use might correlate with health and academic indicators for middle school-aged children in the United States. With that being said, there are a multitude of studies that research the influence of technology use in adolescents and its effect on their physical, mental, interpersonal, academic, and social well-being. In relation to physical health, one study found that children's excessive digital technology use, more specifically, the physical posture of children when using technology, the frequency of using technology, and the duration of using technology of resulted in not only developmental problems but physical problems like physical inactivity, obesity, and sleep disturbances (Mustafaoğlu et al., 2018). Regular physical activity is crucial for children's overall development but that is compromised due to the way technology use encourages a sedentary lifestyle. Mineshita's et al. (2021) study further supports this claim after their results showed a significant association between extensive screen time duration and increased obesity rates and decreased physical activity rates in children. Furthermore, the study also revealed the specific timing of using technology, such as using electronic devices before bedtime, resulted in poorer health outcomes for children. For example, children who watched television or went on their mobile device late at night were prone to late-night eating which not only impacted their weight but also disrupted their sleeping habits. The American Academy of Pediatrics recommends that older children should limit their screen time and technology use to less than two hours per day. Too much screen time can be linked to negative academic performance in addition to negative physical outcomes like obesity. A cross-sectional survey conducted among a sample of Chinese middle school students found that spending more than four hours on social media and electronic devices on both school days and non-school days was negatively associated with academic performance. The study hypothesized that the more hours spent on technology outside of school, the less time students subsequently spent on academic activities like homework and studying for exams (Yan et al., 2017). The amount of time spent using television or electronic devices can shape the way children treat their physical and academic well-being. An excessive amount of time spent on technology outside of school can take away from valuable time spent doing critical development activities like exercise, studying, socializing with peers, and spending time outside and ultimately impact adolescents overall success and well-being.

Another influential component of technology use is the type of content consumed by children throughout the day. A systematic review of television advertisements and childhood obesity found a significant relationship between excessive television use, commercial viewing, and BMI. The review found that many unhealthy food and drink advertisements were played during peak after school hours which may indirectly contribute to children's unhealthy eating habits and behaviors (Kelly et al., 2019). Similarly, various studies have shown that content involving like underage drinking, smoking, and e-cigarette use have a major influence on children's development of these types of maladaptive behaviors. Lee's et al. (2021) study discovered that adolescents that frequently used social media and mobile devices were more susceptible to vape use and did not believe in the harmful effects related to e-cigarette use. More specifically, the study explained that Facebook, Instagram, Twitter, and Snapchat use were associated with vaping in a sample of adolescents. In addition, excessive technology use increases exposure to e-cigarette advertisements on television and social media platforms which has been shown to increase e-cigarette awareness and use among youth (Duan et al., 2021).

Considering the interconnectedness of technology use and other maladaptive behaviors in youth, we also hoped to analyze marijuana use as a closely related maladaptive behavior that affects adolescents well-being. A longitudinal study looking at patterns of marijuana use and physical health among a sample of Canadian youth found significant relationships between marijuana use and reports of negative overall health in youth. The study revealed that the marijuana use can lead to lower engagement in health-promoting behaviors like physical activity and healthy eating habits (Ames et al., 2020). In

addition, there is evidence that marijuana use in adolescence can lead to poorer academic motivation and outcomes as well (Cyrus et al., 2021). One interesting study found that individuals that started using marijuana early in adolescence reported poorer overall health by age 29 (Ellickson et al., 2004). These findings reveal that marijuana use in childhood can impact the kinds of habits and behaviors needed to sustain a healthy life into adulthood. In relation to affecting adolescents' health behaviors, marijuana use has found to be a significant influence on whether adolescents use vape or e-cigarette devices. For example, Park et al. (2020) found that individuals who used e-cigarettes to any degree were also more likely to use marijuana and alcohol. Also, adolescents who frequently smoked marijuana tended to have higher rates of nicotine dependence (Brook et al., 2016). A study focusing on marijuana and nicotine in relation to brain functioning found that marijuana and nicotine users had reduced brain connectivity in connection to non-users (Filbey et al., 2018). These results suggest that marijuana and e-cigarette use may have underlying neurological implications that facilitate excessive use in adolescents. Prior studies show strong relationships between television, electronics, and marijuana use on the various domains of adolescent health and success. Taken all together, our study aims to uncover the effects of maladaptive behaviors such as technology use and drug use on adolescents' well-being.

For our final project, we analyzed data collected by the CDC's Youth Risk Behavior Surveillance System (YRBSS). This is a voluntary survey that was established in 1991 and is conducted biennially by states, territories, tribal governments, and local school districts across the country. The survey codifies and collects hundreds of data points on adolescent behavioral tendencies to drive improvements in youth-related public health policies.

About the Dataset

The raw survey dataset was retrieved from the YRBSS section of the CDC website (https://www.cdc.gov/healthyyouth/data/yrbs/index.htm). It contains over 2 million rows of data from middle school YRBSS surveys conducted from 1995-2019. Due to this magnitude, the dataset proved difficult to work with initially. To overcome its complexity, we reduced the size by focusing our analysis on surveys conducted from 2015-2019. After reviewing all questions, we narrowed in on the following behavioral features:

- **Television use** – hours of TV watched on an average school day
- **Electronics use** – hours of non-school related electronic device use on an average school day
- **Physical activity** – number of days they were physically active for at least 60 minutes within the past week
- **Marijuana use** – whether the respondent has consumed marijuana before
- **Vape use** – whether the respondent has used electronic vapor products before
- **Grades** – description of their grades over the past 12 months

We also utilized the respondent's demographic features of weight, BMI, and race in our models to categorize chosen behaviors.

SMART Questions

Technology has advanced drastically over the past 20 years and has become an integral part of everyday life. Adolescents are often some of the first groups to adopt new technology and use it as a main source of information, entertainment, and communication amongst their peers. With access to technology and social media also comes exposure to risky behaviors such as drug use, which can be detrimental to adolescent development.

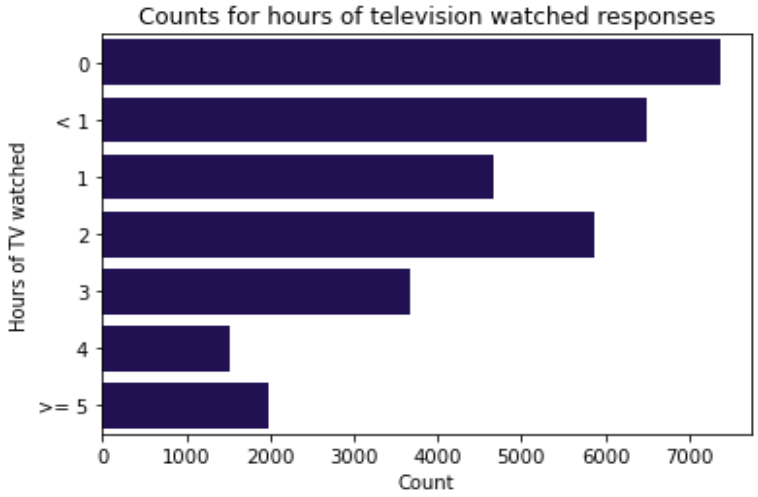
Race and ethnicity also play a key role in helping shape adolescents' environment and how they experience the world. In the United States, cultural resources and family socioeconomic capital expose differences in problem behaviors between Asian American, Black, and Latino adolescents and their White peers. Research has shown that racial/ethnic minority status can also be linked to differences with a variety of markers of adolescent health.

To explore these possibilities, we formulated the following preliminary questions:

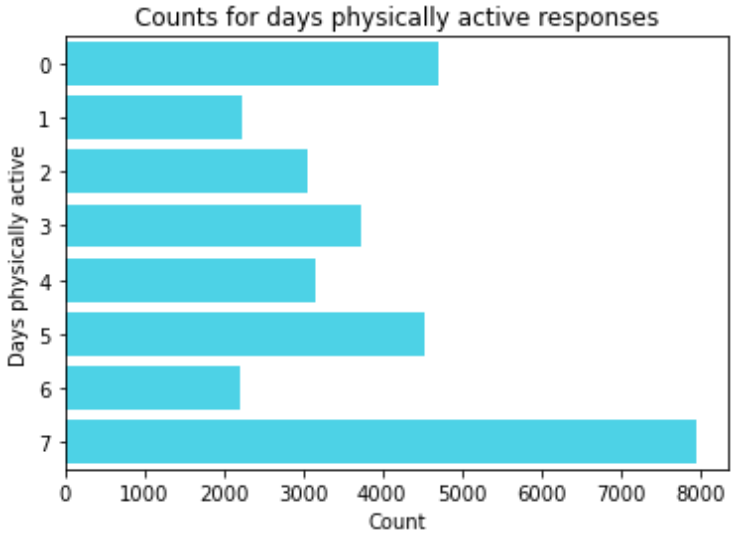
1. **How does technology and drug use relate to positive health and academic outcomes in adolescents?**
2. **Do adolescents of various races differ in their physical health and academic success?**

Exploratory Data Analysis

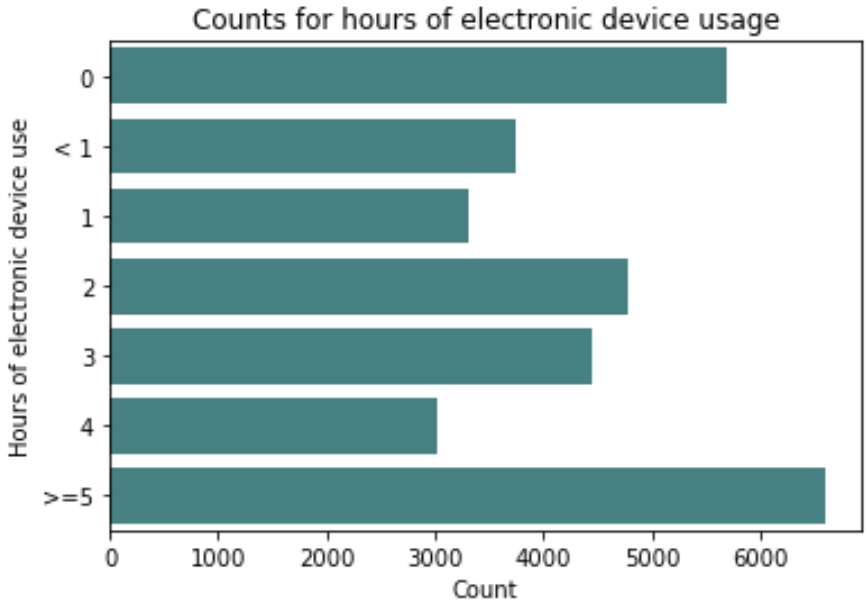
We begin our analysis by visualizing the distribution of responses for our target questions and demographic indicators from the survey.



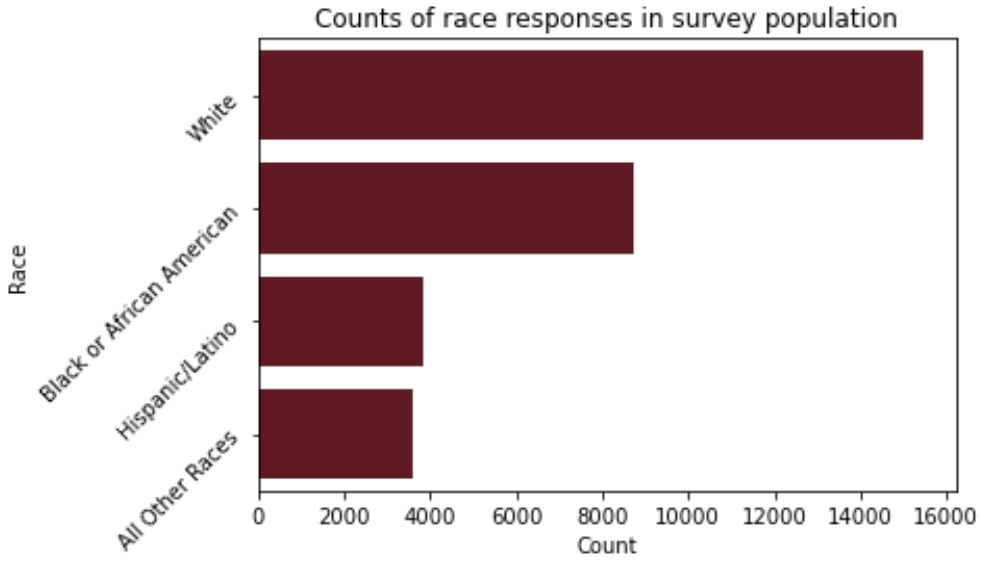
On an average school day, a majority of adolescents responded as watching no TV at all. A large proportion of the survey respondents watched 2 or less hours of television.



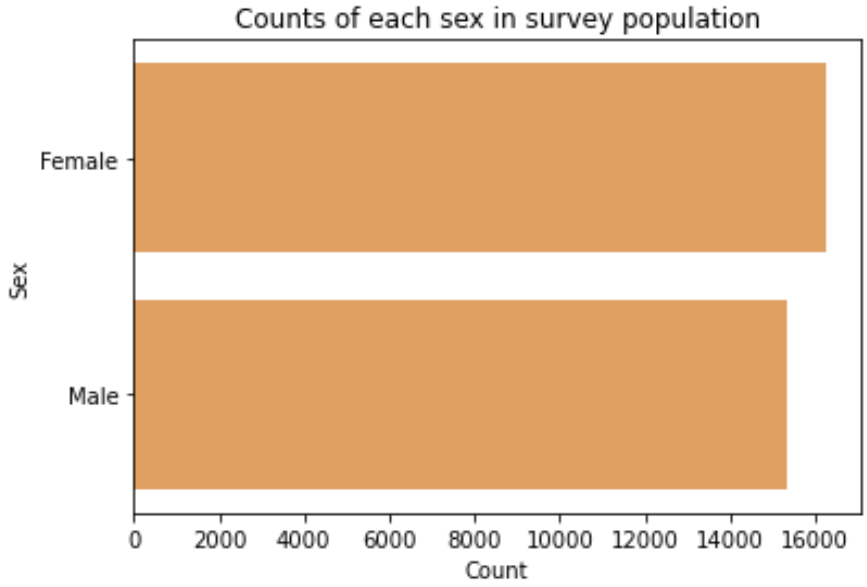
A large majority of adolescents responded as exercising for at least 60 minutes everyday in the past week. The second-most selected response was zero days, although this could be perceived as the adolescent exercising for less than 60 minutes within a given day in the past week.



On an average school day, a majority of adolescents selected using an electronic device for 5 or hours to perform non-school related work. Most adolescents use an electronic device for at least 2 hours, but a larger-than-expected amount of them responded as not performing any non-school related work on an electronic device at all.



A large majority of adolescents identified as White, with Black or African-American as the second most selected choice, and Hispanic/Latino coming in third. Race/ethnicities outside of these 3 groups were lumped together into a single category to simplify analysis. Overall, this is consistent with race proportions observed within the national population.



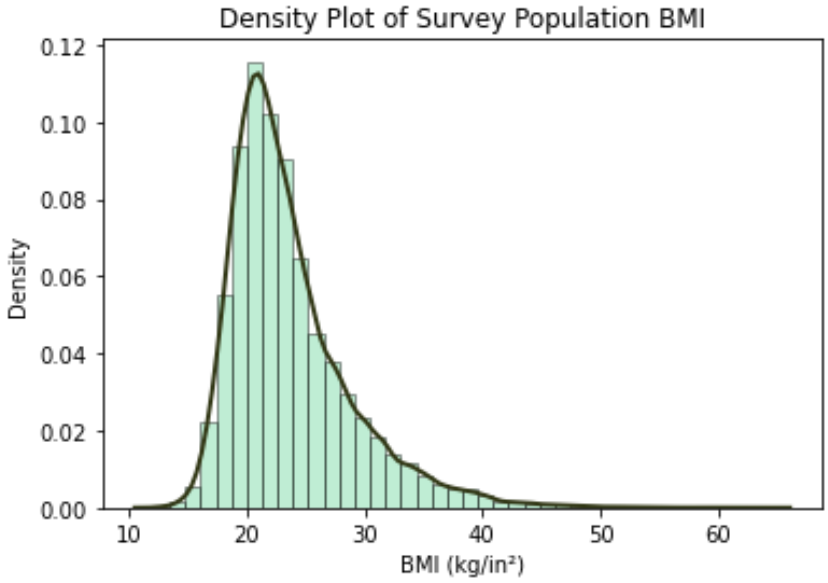
Sex is usually categorized as female or male but there is variation in the biological attributes that comprise sex and how those attributes are expressed. This survey only offers male or female as choices so there is potential for a certain part of the survey population to be misidentified. The distribution between adolescents identifying as either sex is fairly equal, with females having a slight advantage.

Adolescent Body Mass Index

The Body Mass Index (BMI) is a value that is derived from the mass and height of a person. It is defined as the body mass (in kilograms) divided by the square of the body height (in meters), and represented in units of kg/m^2.

In the United States, the CDC utilizes standardized growth charts to monitor the growth of adolescents aged 2 through 19. The CDC growth charts are a national reference that represent how US adolescents grew primarily during the 1970s through 1990s, and instruct health care providers on how to assess physical growth among children and teens.

BMI within our dataset is reported as a raw value which must be interpreted using the age and sex of each adolescent. We believe BMI can be a good indicator of an adolescent's physical health status and are selected to analyze it further.

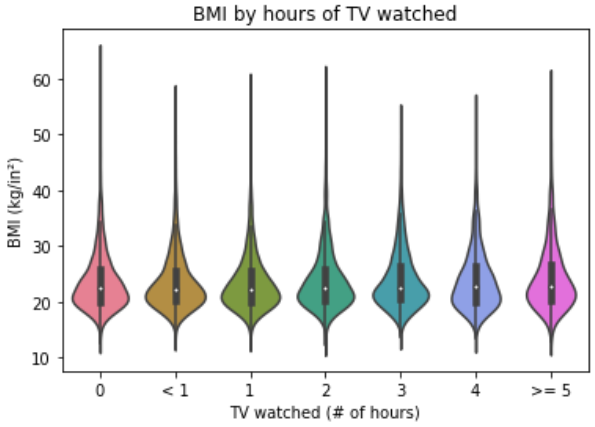


The BMI distribution within the survey population is slightly right-skewed due to uncommonly high BMI outliers. The distribution can be assumed normal based on the plot and large number of observations in the dataset.

Physical Outcomes and Adolescent Behaviors

We elected to analyze the relationship between BMI and adolescent behaviors to better understand the effects of these behaviors on the adolescent's physical health. We can find meaningful correlation factors using plots and statistical tests.

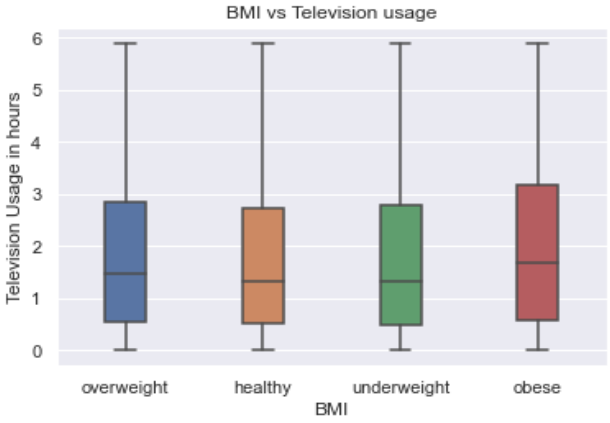
Television Watching & BMI



In the visualization above, we observe slight differences between the average BMI across each response option to the television viewing survey question. We investigated the significance of these differences using a one-way ANOVA test.

F-statistic	13.256
p-value	$4.657e^{-15}$

The p-value of the test confirmed that there are significant differences between the average BMI across groups of adolescents with differing television watching behaviors.

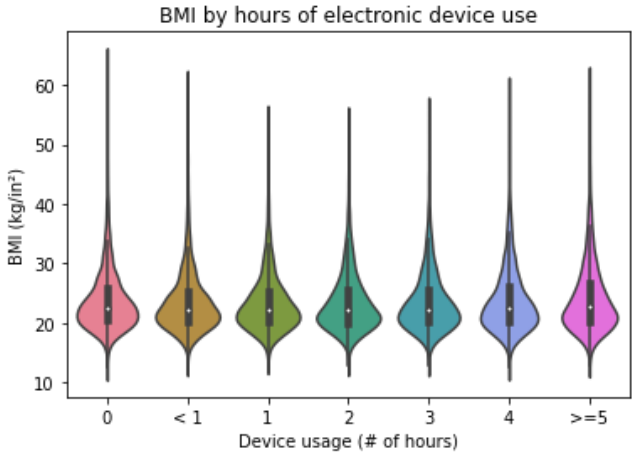


Underweight, healthy, overweight, and obese are the four categories of BMI that have been depicted in the above graphic.According to the graph on the right, obese people watch more television than the other groups—between 0 and 3 hours each week. In contrast, healthy individuals, those who are underweight, and those who are overweight watch television for an average of less than three hours each day.

	Chi-square test	results
0	Pearson Chi-square (18.0) =	103.5175
1	p-value =	0.0000

We can rule out the null hypothesis that BMI and the number of hours spent watching television are independent variables with a p-value that is close to 0. There is evidence to support the relationship between BMI categories and television viewing hours.

Electronic Device Usage & BMI

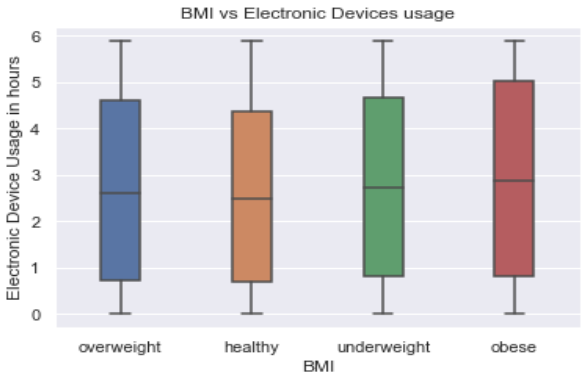


Based on the plot above, we observe differences between the average BMI related to the answer choice for hours of electronic device usage to perform non-school related work.

To verify our observation, we run a one-way ANOVA test.

F-statistic	14.012
p-value	$5.412e^{-16}$

With a p-value close to 0, we can reject the null hypothesis that the average BMI across electronic device usage is equal. There is evidence to suggest that the average BMI is significantly different based on the amount of hours adolescents spent using electronic devices for non-school related work.

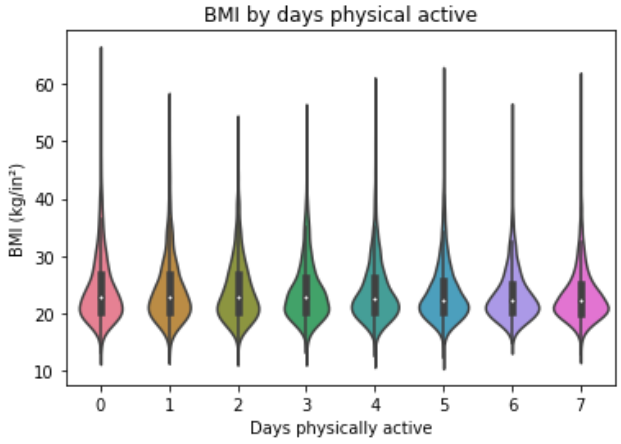


According to the graph, those with underweight BMIs prefer to use electronic devices for longer periods of time, ranging from 1-4.5 hours. In contrast, those with a healthy BMI use electronics at a rate that is nearly identical to that of underweight people. The use of electronic devices by those with overweight body mass index and, lastly, those who are obese tends to exceed 4.5 hours with a maximum daily activity of 5 hours.

	Chi-square test	results
0	Pearson Chi-square (18.0) =	150.3861
1	p-value =	0.0000

We can rule out the null hypothesis that BMI and the number of hours spent using electronic devices are independent variables with a p-value that is close to 0. There is evidence to support the relationship between BMI categories and electronic device usage hours.

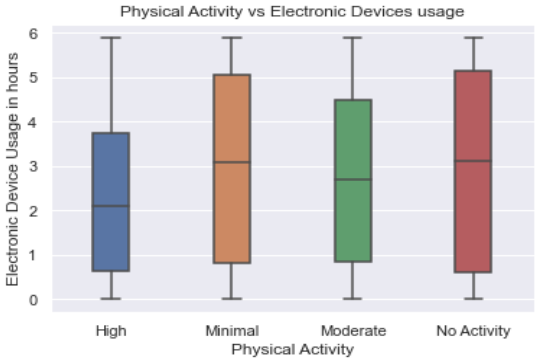
Physical Activity



From the plot above, we are unable to clearly discern differences in BMI distribution across answer choices for the physical activity question. In order to verify, we can conduct a one-way ANOVA test.

F-statistic	28.949
p-value	$4.299e^{-40}$

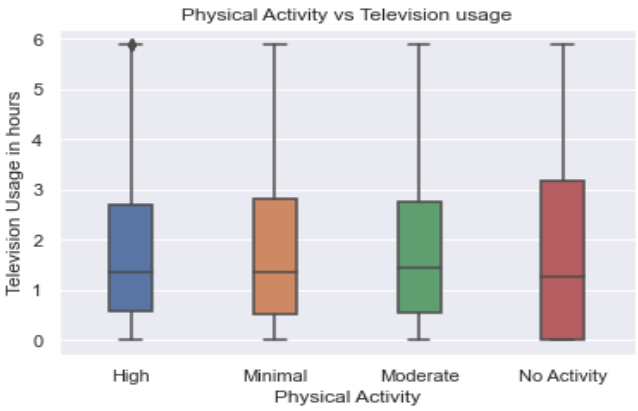
The p-value result of the test confirms the existence of significant differences between the average BMI based on the amount of days adolescents reported being physically active for at least 60 minutes within the past week. This is to be expected since there is a direct, negative correlation between physical activity and weight.



People with No Activity tend to use electronic devices for longer periods of time, spanning from 1 to 5 hours, than the other groups, according to the graph. Electronic gadgets are used almost at the same time by persons who engage in modest physical activity as by those who do not. People with high levels of physical activity tend to use less electronic gadgets, with maximum activity of 3.5 hours, whereas those with moderate levels of activity use them for between one and four hours.

	Chi-square test	results
0	Pearson Chi-square (18.0) =	1133.3054
1	p-value =	0.0000

We can rule out the null hypothesis that Physical Activity and the number of hours spent using electronic devices are independent variables with a p-value that is close to 0. There is evidence to support the relationship between BMI categories and electronic device usage hours.

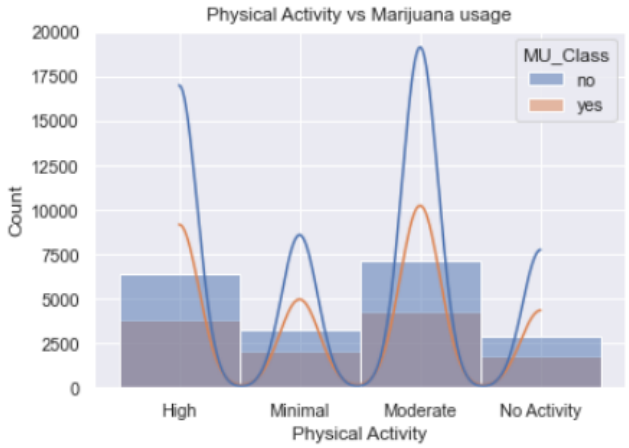


According to the graph on the left, People with No Activity tend to watch more television than the other categories, between 0 and 3 hours per week. While those who engage in high, low, and moderate levels of physical exercise tend to watch TV for fewer than three hours per day on average.

	Chi-square test	results
0	Pearson Chi-square (18.0) =	517.2241
1	p-value =	0.0000

We can rule out the null hypothesis that Physical Activity and the number of hours spent using Television are independent variables with a p-value that is close to 0. There is evidence to support the relationship between BMI categories and electronic device usage hours.

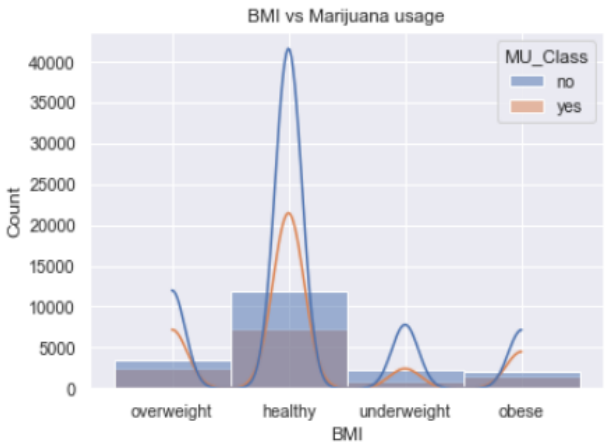
Marijuana Use



Marijuana use and physical activity are shown on the graph. We can see that those who engage in moderate to high levels of physical activity tend to use marijuana more frequently than those who do not. However, the same number of people use marijuana whether they are active or not.

	Chi-square test	results
0	Pearson Chi-square (3.0) =	6.3025
1	p-value =	0.0978

With a p-value larger than 0.05, the null hypothesis that Physical Activity and Marijuana Use are independent variables cannot be ruled out. There is no proof that certain BMI categories and marijuana use are related.



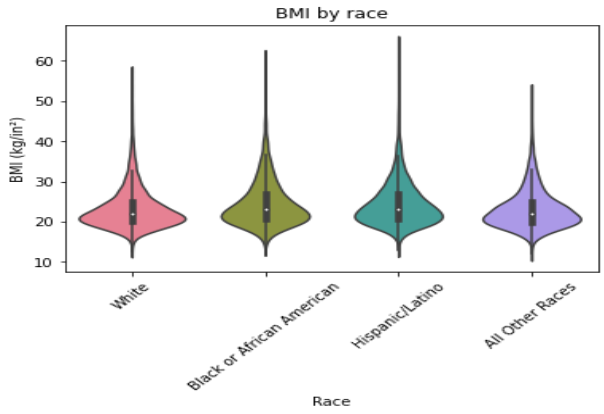
The graph shows marijuana consumption in relation to BMI. We can see that a greater proportion of individuals with healthy BMI tend to use marijuana frequently, followed by individuals with excess body weight. Those who are overweight or obese tend to use marijuana less frequently than those who are normal weight.

	Chi-square test	results
0	Pearson Chi-square (3.0) =	224.4385
1	p-value =	0.0000

We can rule out the null hypothesis that BMI and the marijuana use are independent variables with a p-value that is close to 0. There is evidence to support the relationship between BMI categories and marijuana usage.

Physical Outcomes and Adolescent Demographics

Earlier, we hypothesized that demographic qualities can also lead to different health outcomes for adolescents in the US. To further assess this premise, we can analyze BMI distribution across race and sex identifiers.



We observe significant differences in BMI distribution across the race identity of adolescents.

F-statistic	129.384
p-value	$2.647e^{-83}$

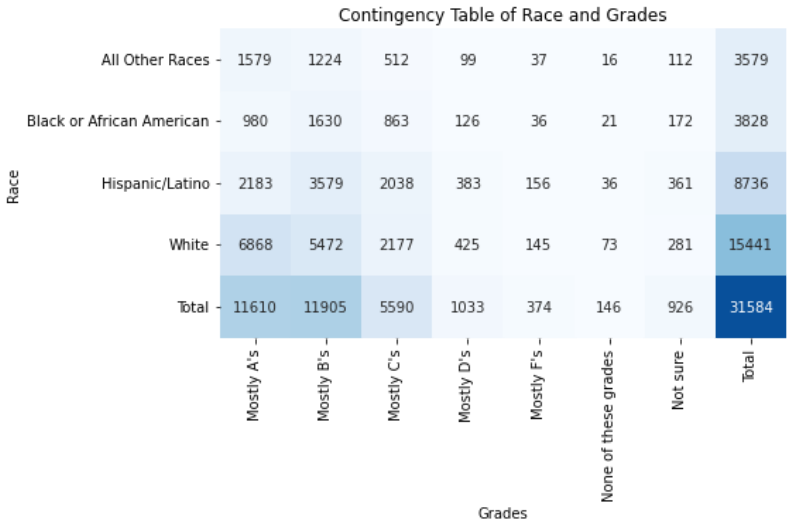
Running a one-way ANOVA results in a p-value near 0, which confirms that average BMI for each of the 4 race choices is significantly different.

The distribution of BMI between each sex appears equal.

F-statistic	13.801
p-value	$2.036e^{-4}$

Running a one-way ANOVA test between partitions of each sex yields a significant p-value of 0.0002. Based on this result, we can conclude that there is a statistically significant difference between average BMI of each sex.

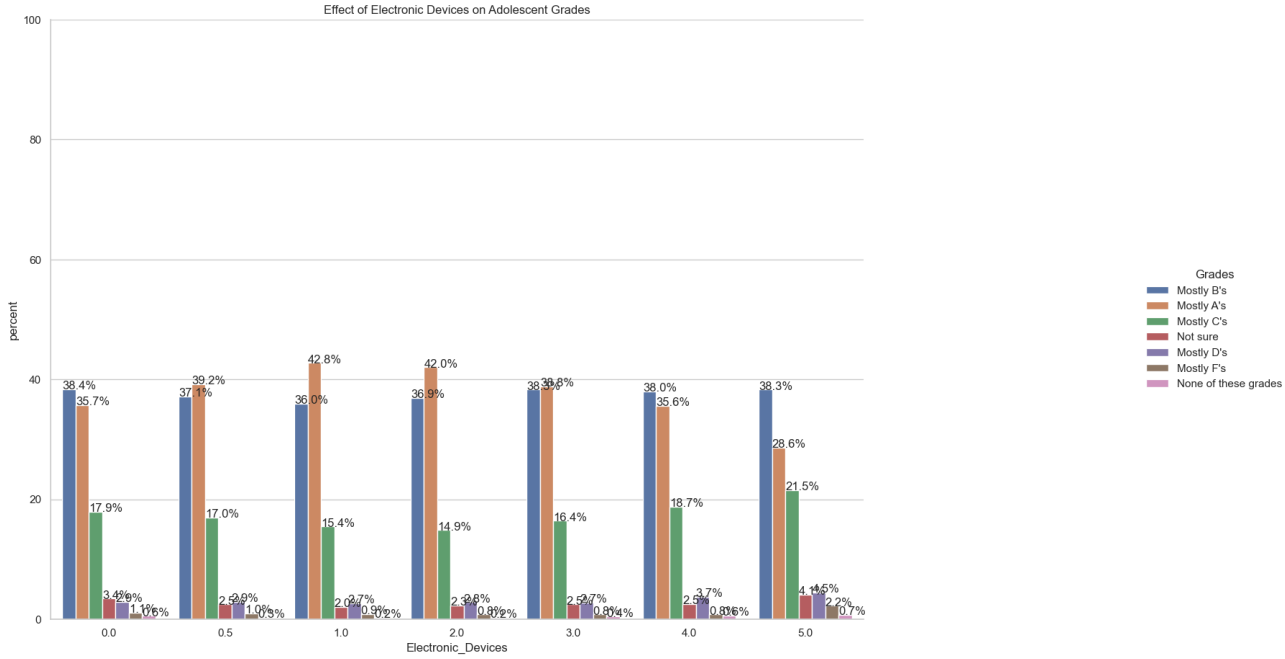
Adolescent Demographics and Grades



The contingency table between racial groups and their grades reveals that a majority of individuals, regardless of race, report having mostly A's and B's for their grades. A majority of white individuals and individuals of other races have mostly A's while a majority of Black/African American and Hispanic/Latino students report having mostly B's. Furthermore, the results of the chi-squared test of independence showed that the p value is less than 0.05 which indicates a significant dependent relationship between race and grades.

Adolescent Behaviors and Grades

Effect of Electronic Devices on Adolescents' Grades



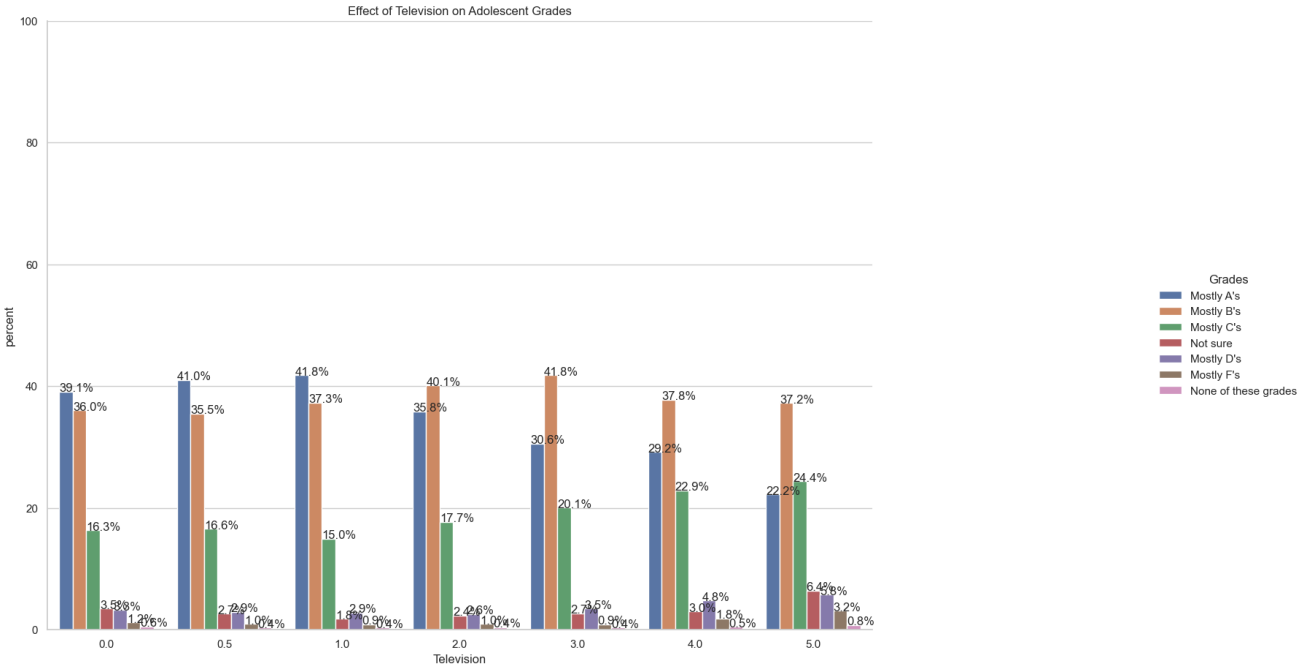
The graph above shows how the use of electronic devices affects adolescents' grades. When adolescents use electronics for two hours, 43% of them get A grades. About 39% of adolescents who use electronics for three hours get B grades and 5 hours of electronic gadget use by adolescents results in C Grades for 22% of them. It appears that there is a correlation between the amount of time spent using electronic devices and grades received. Those who use electronic devices for less than two hours are more likely to receive higher grades, while those who use electronic devices for more than three hours are likely to receive lower grades.

It's important to note that this chart only shows a correlation between electronic device use and grades, and does not necessarily prove that using electronic devices causes lower grades. There could be other factors at play that are contributing to the differences in grades.

It's also worth noting that excessive use of electronic devices may have negative effects on academic performance. Studies have shown that excessive use of electronic devices can interfere with sleep, which is important for learning and cognitive development.

It's important for adolescents to find a balance between electronic device use and other activities that support their physical, mental, and emotional well-being. It's also important for parents, educators, and other adults to provide adolescents with guidance and support to help them make healthy choices about their use of electronic devices.

Effect of Television on Adolescents' Grades

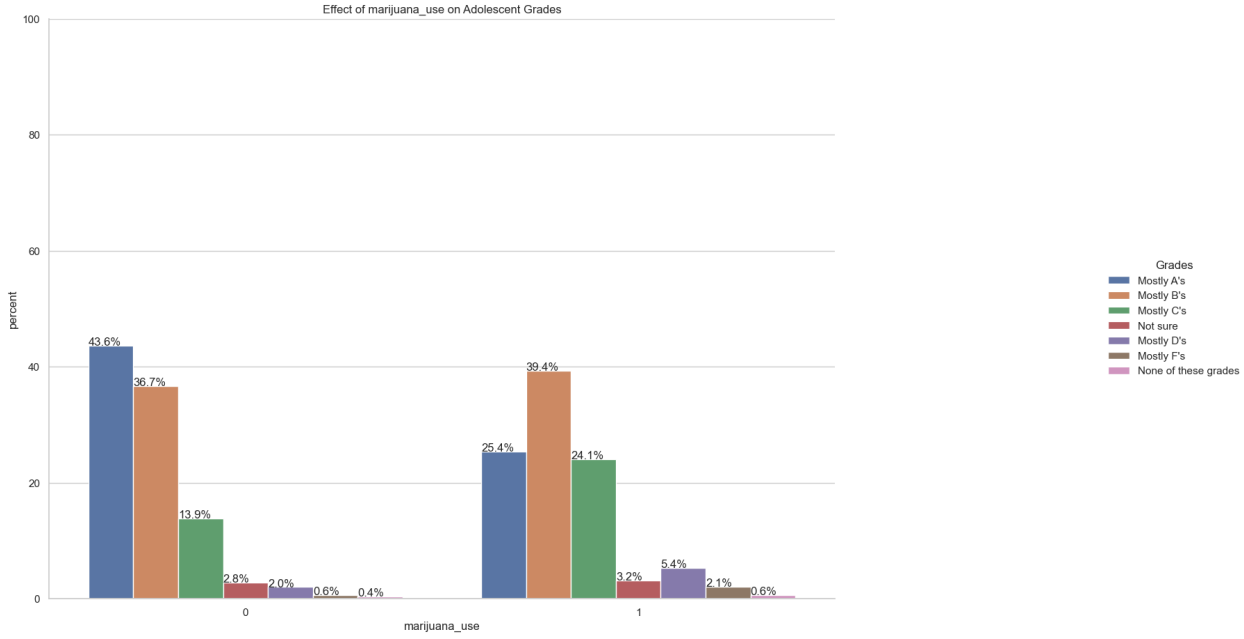


About 41% of adolescents who watch an hour of television earn A grades while another 41% adolescents who watch three hours of television obtain B grades. Adolescents who watch television for 5 hours on average obtain mostly C grades.

It's important to note that this chart only shows a correlation between television viewing and grades, and does not necessarily prove that watching television causes lower grades. There could be other factors at play that are contributing to the differences in grades.

It's also worth noting that excessive television viewing may have negative effects on academic performance. Studies have shown that excessive television viewing can interfere with sleep, which is important for learning and cognitive development. It can also reduce the amount of time that students have available for homework and other activities that support their academic development. It's important for adolescents to find a balance between television viewing and other activities that support their physical, mental, and emotional well-being. It's also important for parents, educators, and other adults to provide adolescents with guidance and support to help them make healthy choices about their media consumption.

Effect of Marijuana on Adolescents' Grades



A grades are earned by 43 percent of adolescents who don't use marijuana.About 24% of adolescents who consume marijuana obtain a C grade, while 38% of them receive a B grade.

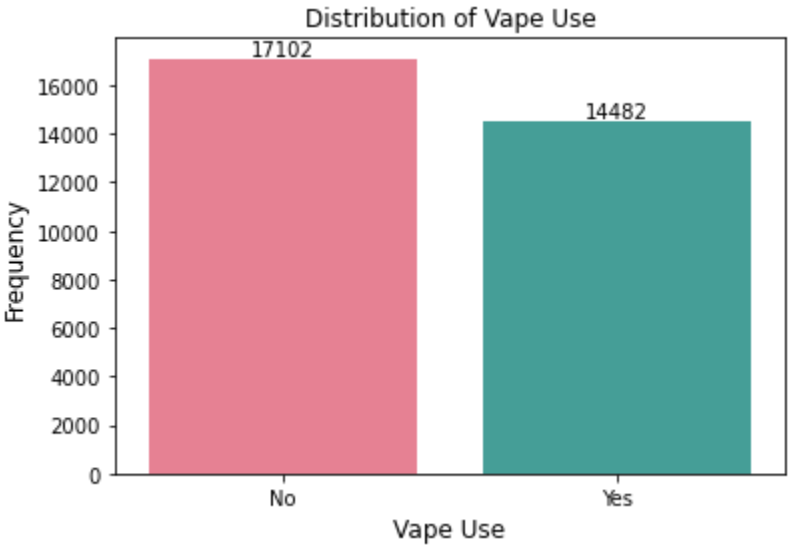
It's important to note that this chart only shows a correlation between marijuana use and grades, and does not necessarily prove that marijuana use causes lower grades. There could be other factors at play that are contributing to the differences in grades.

It's also worth noting that marijuana can have negative effects on cognition and academic performance, so it's possible that students who use marijuana may have lower grades as a result. However, it's also possible that students who have lower grades may be more likely to use marijuana.

It's important for adolescents to be aware of the potential risks and consequences of marijuana use, and to make informed decisions about their own behavior. It's also important for parents, educators, and other adults to provide adolescents with accurate information about marijuana and to support them in making healthy choices.

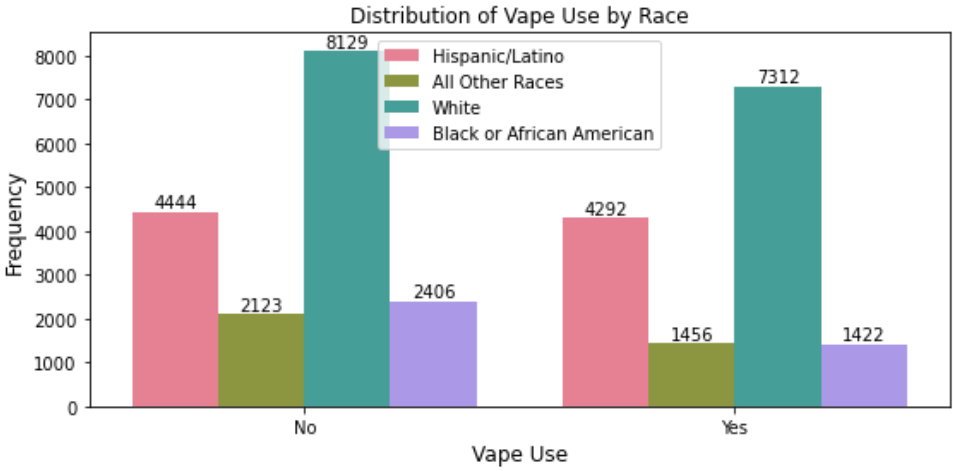
It's important to note that the above charts only show correlations between certain behaviors (TV watching, electronic device use, and marijuana use) and grades, and do not necessarily prove that these behaviors cause lower grades. There could be other factors at play that are contributing to the differences in grades. It's also worth noting that excessive use of electronic devices, marijuana, and television may have negative effects on academic performance. Studies have shown that these behaviors can interfere with sleep, which is important for learning and cognitive development. They can also reduce the amount of time that students have available for homework and other activities that support their academic development. It's important for adolescents to find a balance between these behaviors and other activities that support their physical, mental, and emotional well-being.

Adolescent Behaviors and Vape Use



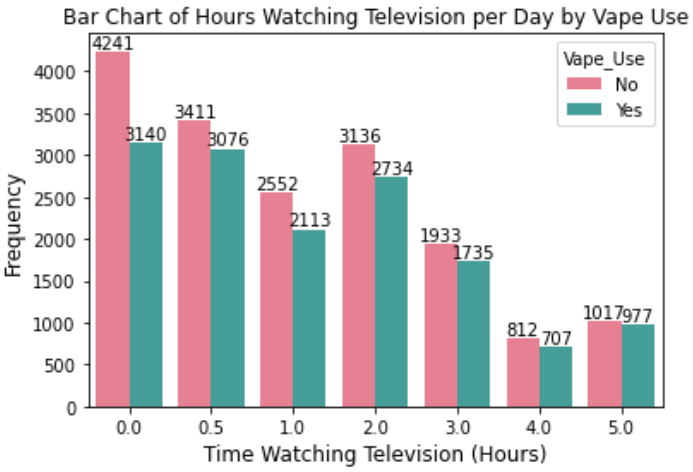
This figure shows the proportion of individuals in the sample who do and do not use electronic vapor products. More specifically, there are 17,102 individuals who do not engage in vaping and 14,482 who do engage in vaping which makes about a 2,620 person difference.

Adolescent Demographics and Vape Use



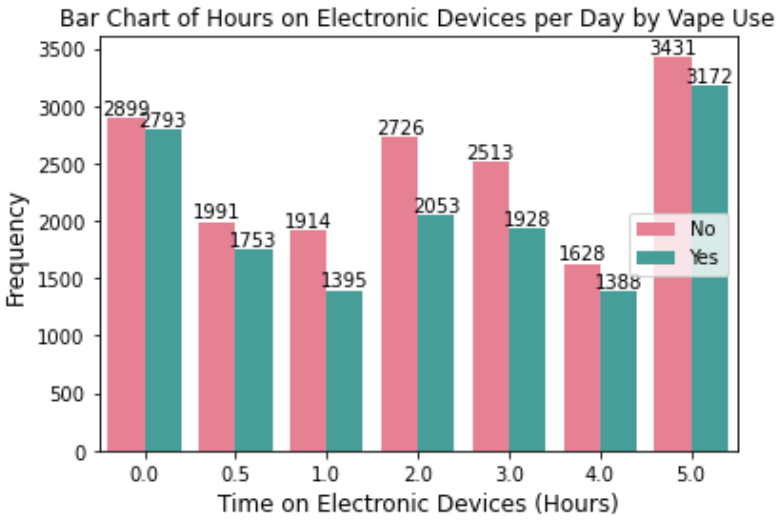
In relation to race and vaping habits, there is a pretty similar distribution between races in terms of individuals that vape and do not vape. More specifically, there is less than a 200 person difference between Hispanic/Latino individuals who vape and do not vape. There is about a 1000 person difference between White individuals, Black/African American individuals, and individuals of all other races who vape and do vape. According to these results, there are definitely differences between racial groups, in comparison to differences within racial groups, in relation to vape use.

Hours of Television and Vape Use



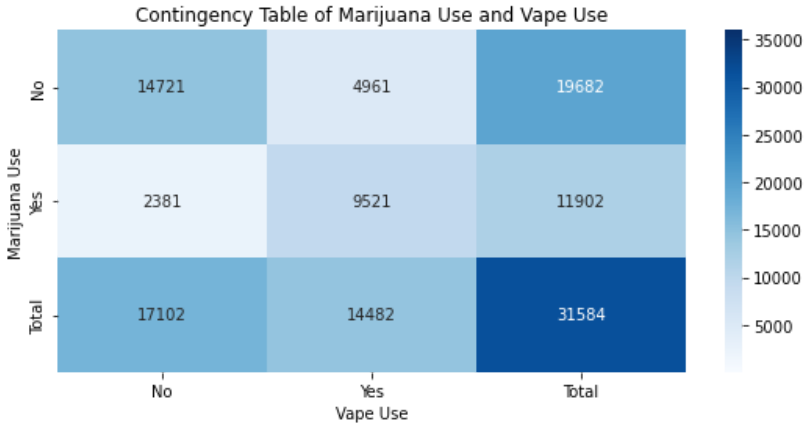
The plot shows the differences in hours of television watched per day by individuals who do and do not vape. Interestingly, between the hours of 0.0 and 2.0 there are many more individuals who report not vaping. In comparison, between the hours of 3.0 and 5.0 it is apparent that a greater proportion of individuals report vaping. It is important to note the pattern that the more hours of television watched in the day, the more the individuals report vaping in comparison to not vaping. These results may imply a relationship between number of hours of television per day and vaping habits considering that the gap between those who vape and those who do not vape becomes smaller and smaller with every extra hour of television watched per day. In addition, after running a two-sample t-test between those who and those who do not vape, the results indicate that there is a significant difference in the average number of hours of television watch per day between groups ($p<0.05$).

Hours of Electronic Device Usage and Vape Use



The plot shows mixed results with the two ends of the hour distribution having the smallest differences between those who do and do not report vaping. The largest difference between groups clusters at 2.0 hours. Overall, there is no definite trend in this graph depicting differences in time spent on electronic devices per day between vaping and non-vaping individuals. This conclusion is further supported by the insignificant ($p>0.05$) t-test which indicates that there is no significant difference in the average number of hours spent on electronic devices between the vaping and non-vaping groups.

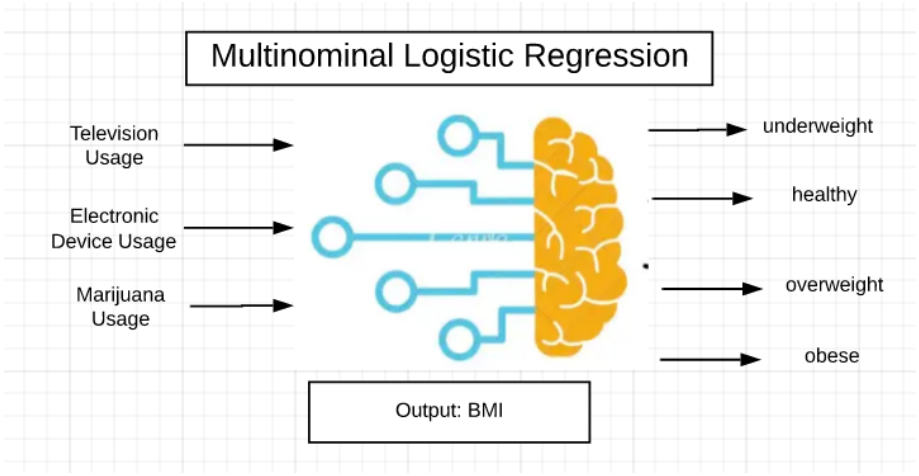
Smoking Marijuana and Vape Use



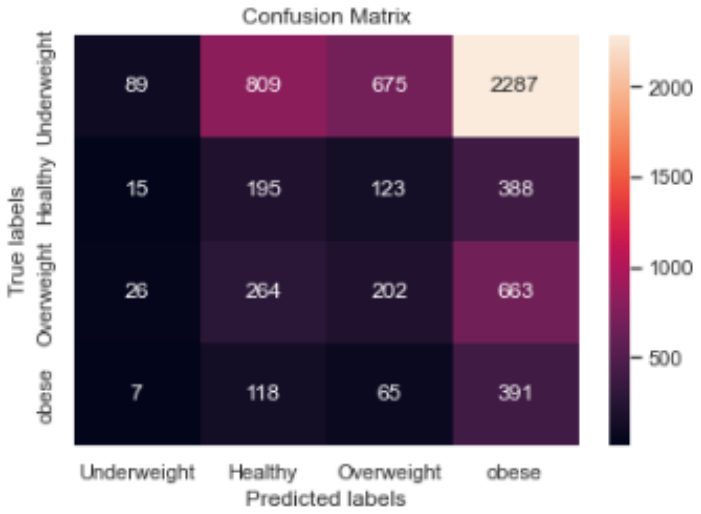
The contingency table between marijuana and vape usage shows that about half of the sample neither vape or smoke marijuana (~46%) while about 30% of the sample did report vaping and smoking marijuana. A very small percentage of the sample either smoke marijuana or vape but do not engage in both. In addition, our chi-squared test of independence indicated a significant dependent relationship between marijuana and vape use.

Models and Model Evaluation

Predicting BMI – Multinomial Logistic Regression

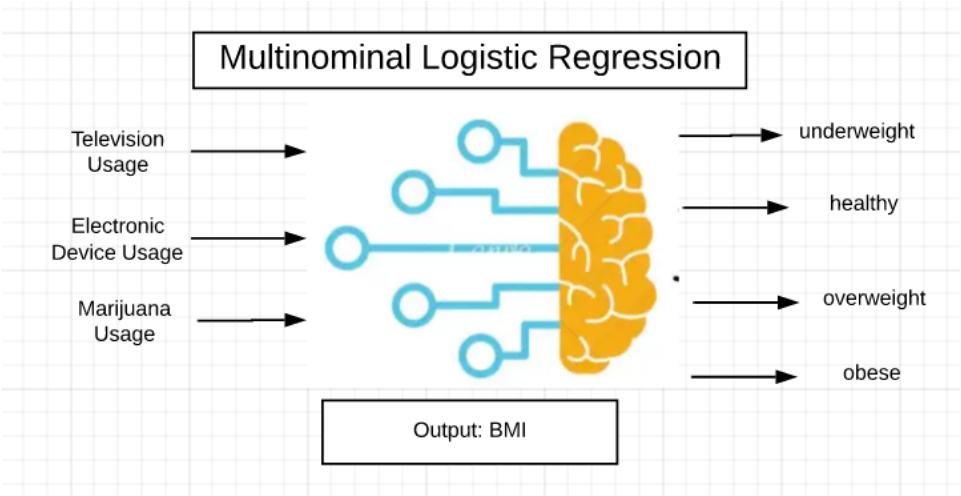


Since the target variable needs to be predicted for numerous categories, we are investigating multinomial logistic regression in this case. The feature variables are television viewing, electronic device use, and marijuana use, and the target variable is BMI.

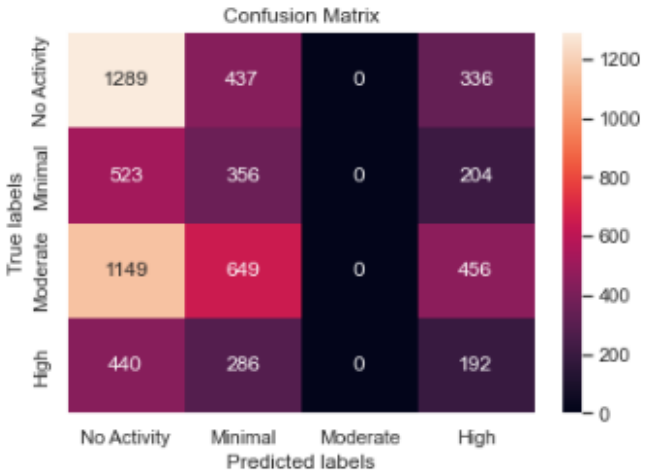


The classification accuracy of the model was found to be 39.54%, while the classification error came out to be 60.46%, indicating that the model's classification accuracy is lower and the adolescents BMI is not classified accurately. However, the precision of the model was found to be around 64.96%, indicating that these many adolescents' BMI were correctly classified by the model, and Recall was 0.0231%. Considering that the pseudo R-Squared Value for the model is 0.065, which indicates that it may or may not correctly identify the BMI of adolescents, the model's overall prediction of 39.54% is not a favorable indicator.

Predicting Physical Activity – Multinomial Logistic Regression



Since the target variable needs to be predicted for numerous categories, we are investigating multinomial logistic regression in this case. The feature variables are television viewing, electronic device use, and marijuana use, and the target variable is physical activity.



Our Multinomial Logistic Regression model's Confusion Matrix shows that its Classification Accuracy was 54.33% and its Classification Error was 45.67%. This indicates that the model accurately identified the presence of physical activity in more than 50% of the adolescents in the dataset. While recall is 62.51% and precision is around 37.90%, the model appears to have accurately categorized these many teens. When you consider the pseudo R-Squared Value of 0.0172, which indicates that the model may or may not correctly classify adolescents physical activity, the overall model prediction of 54.33% is not a positive indicator.

Predicting Grades – Decision Tree Classifier


```
DecisionTreeClassifier TEST:  0.5820435112025111
```

In this project, we used a decision tree to estimate grades based on important variables such as race, television viewing, electronic device use, and marijuana use. The goal was to understand the relationship between these variables and grades, and to use this information to make predictions about grades. The accuracy of the decision tree was not very high, at only 58%. To build the decision tree, we first collected data on these variables for a sample of adolescents. We then used the Synthetic Minority Oversampling Technique (SMOTE) to balance the class distribution in the dataset. The resulting synthetic samples were added to the original dataset, which was then used to train the decision tree. One possible reason for the low accuracy of the decision tree could be the complexity of the problem. Grades are likely influenced by a wide range of factors, including academic ability, study habits, and personal circumstances, which may be difficult to capture with a decision tree. Additionally, the sample size of the study may not have been diverse enough to accurately reflect the relationship between the variables and the outcome.Despite the low accuracy of the decision tree, the variables included in the model can still be useful for understanding the factors that influence grades. Overall, the results of this study suggest that it may be challenging to accurately predict grades based on the variables considered in this analysis. Further research is needed to better understand the factors that contribute to academic performance and to identify effective strategies for improving grades.

Predicting Vape Use – Logistic Regression Model

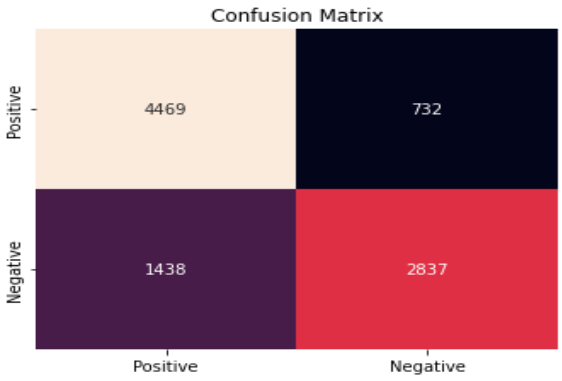
Generalized Linear Model Regression Results

Dep. Variable:	Vape_Use	No. Observations:	31584
Model:	GLM	Df Residuals:	31577
Model Family:	Binomial	Df Model:	6
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-16859.
Date:	Sun, 18 Dec 2022	Deviance:	33719.
Time:	03:14:17	Pearson chi2:	3.17e+04
No. Iterations:	5	Pseudo R-squ. (CS):	0.2679
Covariance Type:	nonrobust		

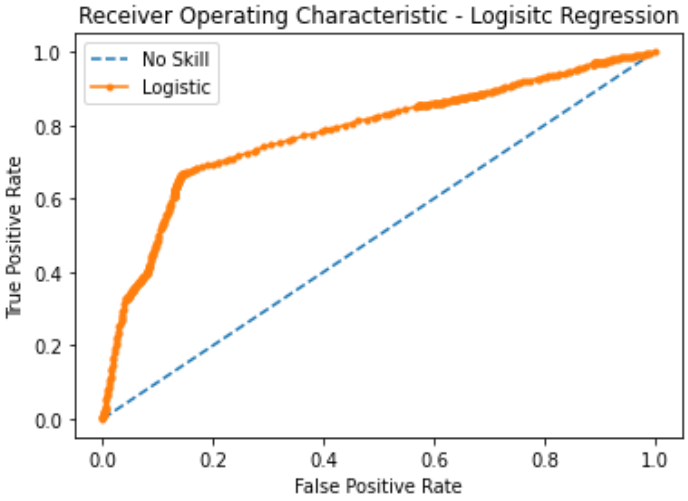
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.9495	0.028	-33.900	0.000	-1.004	-0.895
C(marijuana_use)[T.1]	2.5433	0.029	87.479	0.000	2.486	2.600
C(race)[T.1]	-0.9010	0.046	-19.762	0.000	-0.990	-0.812
C(race)[T.2]	-0.1702	0.032	-5.294	0.000	-0.233	-0.107
C(race)[T.3]	-0.2988	0.045	-6.704	0.000	-0.386	-0.211
Television	0.0260	0.009	2.774	0.006	0.008	0.044
Electronic_Devices	-0.0064	0.007	-0.863	0.388	-0.021	0.008

Logit model accuracy (with the test set): 0.7710004221190375
Logit model accuracy (with the train set): 0.7660575357336711

We choose to run a logistic regression model in order to classify vape use due to its effectiveness and ability to interpret coefficients as a way to tell which predictors are most impactful for our model of vape use. After converting the coefficients to their exponential form, it can be said that in comparison to individuals that do not smoke marijuana, the odds-ratio of using vape products is multiplied by 12.7 for individuals who do smoke marijuana, when holding all other predictors constant. Also, holding all other predictors constant, for African American individuals the odds-ratio of using vape products is multiplied by 2.46, multiplied by 1.19 for Hispanic/Lation individuals, and multiplied by 1.35 for individuals of all other races in comparison to White individuals. Lastly, for every hour increase in television use per day, the odds-ratio of using vape products is multiplied by 1.03. The relationship between electronic device usage and vape use was not significant. From our model, we can successfully predict whether adolescents use vape products based on marijuana use, race, and hours of television watched per day.



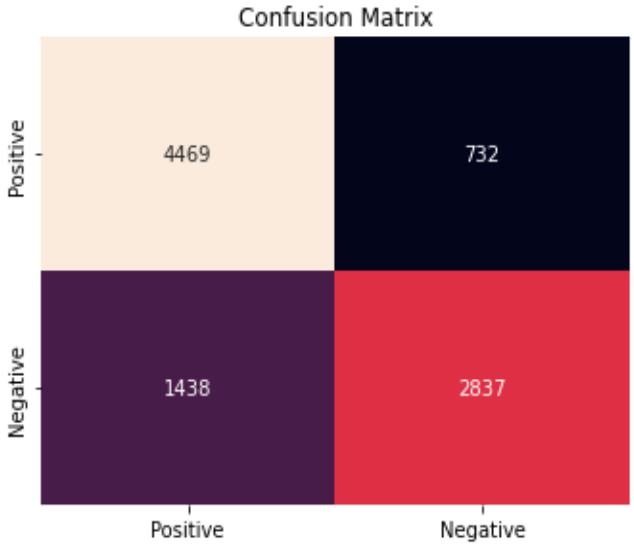
According to the classification report of our logistic regression model, out of all adolescents that the model predicted would use vape products, only about 79% actually do use vape products (precision). Out of all the adolescents that actually do vape, the model only predicted this outcome correctly for 66% of those adolescents (recall). Since the F1-Score is somewhat close to 1, we can assume that the model does a good job of predicting whether or not adolescents will use vape products. The overall accuracy of the model was 77% which is a good sign that the model is efficient at classifying between adolescents who vape and those who do not vape.



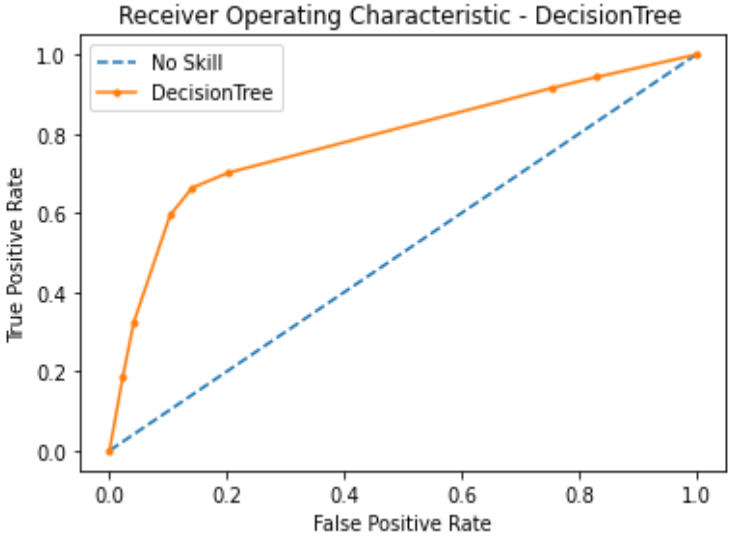
The ROC curve and AUC score help us to understand separability. More specifically, it tells us how much our model is capable of distinguishing between adolescents who do and do not vape. The ROC curve trends somewhat to the upper left corner which indicates a pretty good model. The AUC score is about 0.77 which is acceptable but it would be preferred to be closer to 0.8. Overall, the ROC curve and AUC score indicate that our logistic regression does a good job at discriminating between classes.

Vape Use – Decision Tree Classifier

In order to compare the accuracy of models, we ran a decision tree classification to predict vape use. In comparison to logistic regression, the decision tree classifier had similar accuracy which was around 77%. Our results indicated that the decision tree classifier was able to successfully classify adolescents who do and do not use vape products based on race, marijuana usage, and hours of television watched per day. Marijuana use appeared to be the most important feature for our model and had the greatest impact on predicting vape use.



Similar to our logistic regression model classifying vape use, our decision tree predicted that out of all adolescents that the model predicted would use vape products, only about 79% actually do use vape products (precision). Out of all the adolescents that actually do vape, the model only predicted this outcome correctly for 66% of those adolescents (recall). Since the F1-Score is somewhat close to 1, we can assume that the model does a good job of predicting whether or not adolescents will use vape products. The overall accuracy of the model was 77% which is a good sign that the model is efficient at classifying between adolescents who vape and those who do not vape.



The ROC curve and AUC score tells us how much our model is capable of distinguishing between adolescents who do and do not vape. Like the logistic regression model, the ROC curve trends somewhat to the upper left corner which indicates a pretty good model. The AUC score is about 0.78 which is acceptable and a tiny bit better than our logistic regression but it would be preferred to be closer to 0.8. Overall, the ROC curve and AUC score indicate that our decision tree does a good job at discriminating between classes of vaping and not vaping.

Conclusion

To respond to our smart questions, we could see that there were statistically significant correlations between the variables and the results of the tests of variance (ANOVA) and independence (chi-squared). The logistic regression models used to categorize adolescent grades and health outcomes were unable to produce results with a significantly high level of accuracy due to weak correlations between the variables, but the model used to categorize e-cigarette use produced significant results with a moderately high level of accuracy. Overall, we can see that due to the weak link between the target and feature, marijuana use, television viewing, and electronic device use may or may not have an effect on adolescent physical activity, BMI, and grades. But it is undeniable that vaping is bad for teenagers' health.The use of drugs and technology has an effect on adolescents' health behaviors, according to the overall findings.

References

Ames, M. E., Leadbeater, B. J., Merrin, G. J., & Thompson, K. (2018). Patterns of marijuana use and physical health indicators among Canadian youth. *International Journal of Psychology*, *55*(1), 1–12.https://doi.org/10.1002/ijop.12549

Brook, J. S., Zhang, C., Leukefeld, C. G., & Brook, D. W. (2016). Marijuana use from adolescence to adulthood: developmental trajectories and their outcomes. *Social Psychiatry and Psychiatric Epidemiology*, *51*(10), 1405–1415. https://doi.org/10.1007/s00127-016-1229-0

CDC. National Center for Chronic Disease Prevention and Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition, Activity, P., & Obesity. (2013). *Use and Interpretation of the WHO and CDC Growth Charts for Children from Birth to 20 Years in the United States CDC Recommendation*. https://www.cdc.gov/nccdphp/dnpa/growthcharts/resources/growthchart.pdf

Centers for Disease Control and Prevention. (2019). *Body Mass Index (BMI)*. Centers for Disease Control and Prevention. https://www.cdc.gov/healthyweight/assessing/bmi/index.htm

Cyrus, E., Coudray, M. S., Kiplagat, S., Mariano, Y., Noel, I., Galea, J. T., Hadley, D., Dévieux, J. G., & Wagner, E. (2021). A review investigating the relationship between cannabis use and adolescent cognitive functioning. *Current Opinion in Psychology*, *38*, 38–48. https://doi.org/10.1016/j.copsyc.2020.07.006

Duan, Z., Wang, Y., Emery, S. L., Chaloupka, F. J., Kim, Y., & Huang, J. (2021). Exposure to e-cigarette TV advertisements among U.S. youth and adults, 2013–2019. *PLOS ONE*, *16*(5), e0251203.https://doi.org/10.1371/journal.pone.0251203

Eriksson, I., Undén, A.-L., & Elofsson, S. (2001). Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology*, *30*(2), 326–333. https://doi.org/10.1093/ije/30.2.326

Filbey, F. M., Gohel, S., Prashad, S., & Biswal, B. B. (2018). Differential associations of combined vs. isolated cannabis and nicotine on brain resting state networks. *Brain Structure and Function*, *223*(7), 3317–3326. https://doi.org/10.1007/s00429-018-1690-5

Kann, L. (2001). The Youth Risk Behavior Surveillance System: Measuring Health-risk Behaviors. *American Journal of Health Behavior*, *25*(3), 272–277. https://doi.org/10.5993/ajhb.25.3.14

Kelly, B., Vandevijvere, S., Ng, S., Adams, J., Allemandi, L., Bahena‐Espina, L., Barquera, S., Boyland, E., Calleja, P., Carmona‐Garcés, I. C., Castronuovo, L., Cauchi, D., Correa, T., Corvalán, C., Cosenza‐Quintana, E. L., Fernández‐Escobar, C., González‐Zapata, L. I., Halford, J., Jaichuen, N., & Jensen, M. L. (2019). Global benchmarking of children's exposure to television advertising of unhealthy foods and beverages across 22 countries. *Obesity Reviews*, *20*(S2), 116–128.https://doi.org/10.1111/obr.12840

Lee, J., Tan, A. S. L., Porter, L., Young-Wolff, K. C., Carter-Harris, L., & Salloum, R. G. (2021). Association Between Social Media Use and Vaping Among Florida Adolescents, 2019. *Preventing Chronic Disease*, *18*(18). https://doi.org/10.5888/pcd18.200550

Yan, H., Zhang, R., Oniffrey, T., Chen, G., Wang, Y., Wu, Y., Zhang, X., Wang, Q., Ma, L., Li, R., & Moore, J. (2017). Associations among Screen Time and Unhealthy Behaviors, Academic Performance, and Well-Being in Chinese Adolescents. *International Journal of Environmental Research and Public Health*, *14*(6), 596. https://doi.org/10.3390/ijerph14060596

Zilanawala, A., Bécares, L., & Benner, A. (2019). Race/ethnic inequalities in early adolescent development in the United Kingdom and United States. *Demographic Research*, *40*(6), 121–154.https://doi.org/10.4054/demres.2019.40.6