HEALTH DATA SCIENCE

The Impact of Technology Use on Academic Performance and Life Satisfaction Among Youth

DATE :02 May 2024

Presented By:

Vudem Shruthi Reddy

BACKGROUND INFORMATION

Our research question is important because technology use among children and adolescents has increased dramatically in recent years, yet the impacts on academic performance and life satisfaction are not fully understood. Examining these relationships could provide valuable insights to guide parenting, education, and health recommendations around childrens tech use.

Yes, there have been exiting studies like our research question some of the studies cited by Lepp et al. (2014), Moksnes et al. (2016), Crede et al. (2015), Żerebecki & Opree (2022), and Sampasa-Kanyinga et al. (2022). These studies have found mixed results on the relationships between technology use, academic achievement, and life satisfaction in young people[1,2,3,4,5].

Our project could differentiate itself by focusing specifically on the GSS dataset and the unique SDOH variables it contains. Most previous studies appear to have used other data sources. We examine which specific SDOH factors moderate the relationships between tech use, academics, and life satisfaction.

We thought to use the General Social Survey (GSS) dataset, which is a high-quality data source containing information on demographics, behaviors, and attitudes from a representative US sample. It includes SDOH variables which makes it suitable for this health-related research question.

The dependent variables could be academic performance and life satisfaction i.e EDUC variable and Happy variable. The independent variables could include technology use (USETECH), work status (WRKSTAT), income (RINCOME), marital status (MARITAL), sex, and race. These SDOH factors may influence the relationships between tech use, academics compared to other studies. Relationships between this variables can also examined.

The study's significance lies in its use of the GSS dataset, focus on SDOH factors and inform evidence-based guidelines for healthy technology use among youth. Practical implications could include recommendations for parents, educators, health providers, and policymakers to promote academic success and wellbeing in an increasingly digital world.

METHODOLOGY

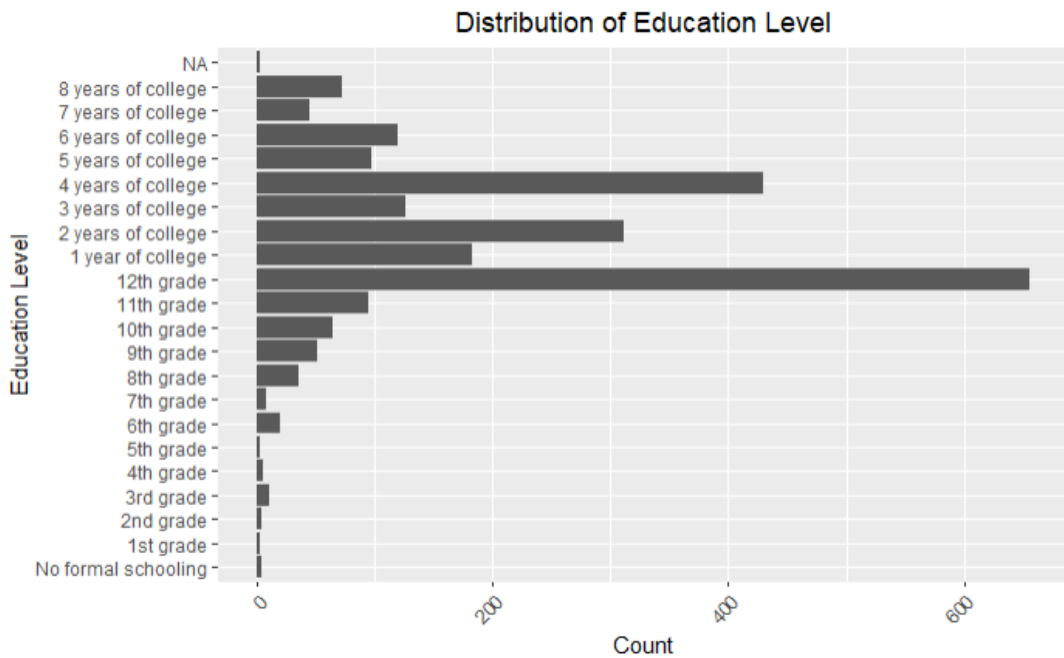
DESCRIPTIVE STATISTICS

	Level	Overall
N		2345
EDUC (%)	No Formal Schooling	4(0.2)
	1 St Grade	2 (0.1)
	2nd Grade	4 (0.2)
	3 Rd Grade	10 (0.4)
	4 th Grade	5 (0.2)
	5 th Grade	3 (0.1)
	6 th Grade	20 (0.9)
	7 th Grade	8 (0.3)
	8 th Grade	35 (1.5)
	9 th Grade	51 (2.2)
	10 th Grade	65 (2.8)
	11 th Grade	95 (4.1)
	12 th Grade	656 (28.0)
	1 year of college	183 (7.8)
	2 years of college	312 (13.3)
	3 years of college	126 (5.4)
	4 years of college	430 (18.4)
	5 years of college	97 (4.1)
	6 years of college	119 (5.1)
	7 years of college	45 (1.9)
	8 years of college	72 (3.1)
Happy (%)	Very Happy	701 (29.9)
	Pretty Happy	1304 (55.7)
	Not Too Happy	336 (14.4)
USETECH (mean (SD))		55.15 (29.32)
Age (mean (SD))		49.13 (18.24)
SEX (%)	Male	1051 (44.8)
	Female	1294 (55.2)
Race (%)	White	1692 (72.2)
	Black	383 (16.3)

	Others	270 (11.5)
WRKSTAT (%)	Working Full Time	1133 (48.3)
	Working Part Time	258 (11.0)
	With a job not working	52 (2.2)
	Unemployed	84 (3.6)
	Retired	445 (19.0)
	In School	81 (3.5)
	Keeping House	242 (10.3)
	Other	48 (2.0)
	No Answer	2 (0.1)
RINCOME (%)	Under $1000	33 (2.3)
	$1000 - $2999	32 (2.3)
	$3000 - $3999	32 (2.3)
	$4000 - $4999	21 (1.5)
	$5000 - $5999	21 (1.5)
	$6000 - $6999	12 (0.8)
	$7000 - $7999	18 (1.3)
	$8000 - $9999	33 (2.3)
	$10000 - $14999	93(6.5)
	$15000 - $19000	61 (4.3)
	$20000 - $24999	107 (7.5)
	$25000 and over	849 (59.8)
	Refused	108 (7.6)
Marital (%)	Married	998 (42.6)
	Widowed	200 (8.5)
	Divorced	402 (17.2)
	Seperated	75 (3.2)
	Never Marrried	668(28.5)
HRS 1(%)	10-19 hours	7 (17.1)
	20-29 hours	2 (4.9)
	30-39 hours	5 (12.2)
	50-59 hours	2 (4.9)
	60-69 hours	8 (19.5)
	70-79 hours	2 (4.9)
	80 and more	6 (14.6)
	No answer	9 (22.0)
HRS 2 (%)	60-69 hours	1(33.3)
	Don’t Know	1(33.3)
	No answer	1 (33.3)

Data Visualization

Distribution of Education Level



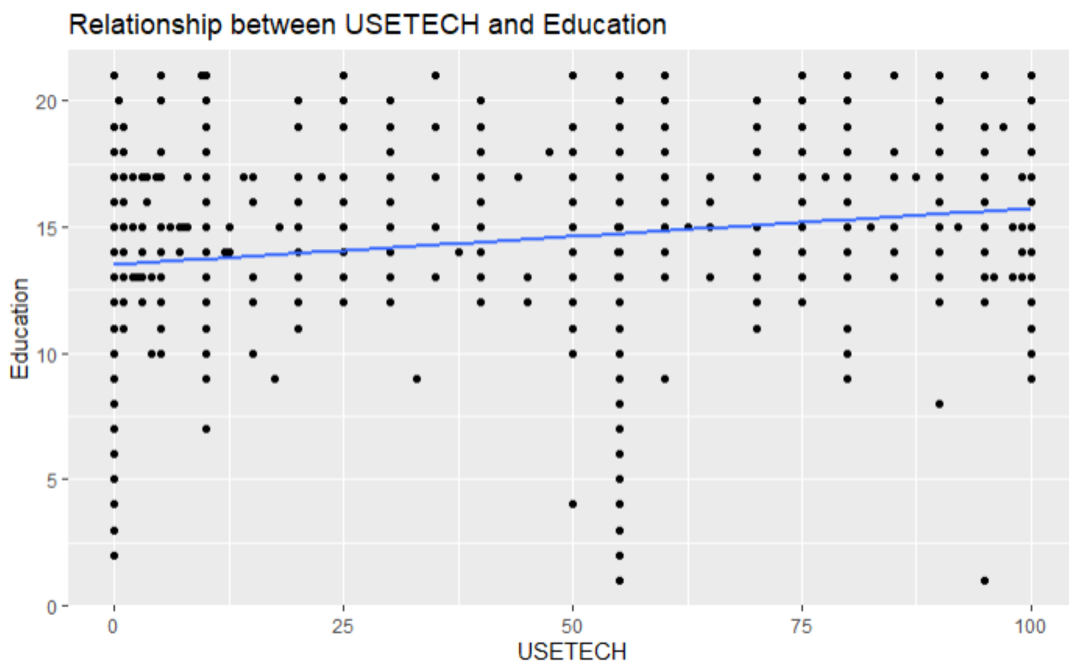
This bar chart shows the division of participants into various education levels. The x-axis is displayed as different education level categories, starting from "No formal schooling" and ending at "8 years of college." The y-axis represents the number of participants in each education level category. The graph shows that the most common level of education is "high school ", then "4 years of university" and "2 years of university". This graph clearly demonstrates the educational attainment of the survey population.

Distribution of Happiness Level

This bar chart shows the distribution of the happiness self-ratings made by the participants. The x-axis shows the three categories of happiness: "Very Happy", "Pretty happy", and "Not Too happy" are depicted on the x-axis, which also shows the count or frequency of the participants in each category. A glance at the chart can be seen that the majority of the people voted for "Pretty Happy," and then "Very Happy," while a smaller number selected "Not Too Happy." This makes it evident that the general happiness level of the sample is high.

Relationship between USETECH and Education

Such a scatter plot study is about the interaction between the participants' technological usage (USETECH) and their educational level. The USETECH variable is represented on the x-axis of the chart, and the y-axis shows the level of education. On the plot, every data point corresponds to the USETECH score of a participant and the education level that they have achieved. The regression line in blue shows the slope of the relationship between the two variables. This visualization is then used to analyze the possibility of the existence of the correlation between the use of technologies and the educational achievement in the given sample.



Data Management

Firstly, the data missing in the “USETECH” variable, coded as -1, 998, and 999, was replaced with NA (missing) using the na_if function in R. Subsequently, the mean value of the non-missing “USETECH” scores was calculated and imputed for the missing values.

Numerous categorical variables were changed from numerical values to descriptor for easier understanding. The "EDUC" variable was recoded from numeric values (0-20) corresponding to education levels (e.g., "No formal schooling", "1st grade"...,"8 years of college") by factor function. Apart from that, "HAPPY", "WRKSTAT", "RINCOME", "MARITAL", "SEX", "RACE", "HRS1", and "HRS2" were assigned descriptive labels.

Based on the research questions and variables of interest, the data set was subset to include only the relevant variables: "EDUC", "HAPPY", "USETECH", "AGE", "SEX", "RACE", "WRKSTAT", "RINCOME", "MARITAL", and "HRS1" & "HRS2".

Moreover, a new variable "USETECH_group" was added and it was divided into "High" and "Low" groups according to the median value of "USETECH", which was used to arrange the participants into different categories. The purpose of this step was to simplify the comparison between groups by using standard statistical techniques like t-tests or ANOVA. Other than that, a new binary variable, "LifeSat", was formed by deriving it from the "HAPPY" variable, having a value of 1 for "Very Happy" and 0 for other happiness levels, which allowed us to use logistic regression to analyze life satisfaction.

Statistical Tests

In order to answer the research questions, different statistical methods were used. To begin with, the mean score differences between the two groups (high USETECH and low USETECH) were determined via independent samples t-test. The t-test is based on the mean and the variance. The Shapiro-Wilk test was employed to verify the normality assumption, while Levene's test was used to evaluate equality of variances.

Since the normal assumption was violated as far as HAPPY variable concerned, the nonparametric Wilcoxon rank-sum test was used as an alternative to compare the HAPPY scores between low and high USETECH groups. As no standardized residuals or post-hoc tests were available for this sort of group comparison, the results were presented as raw mean values.

The two-sample t-test for EDUC in the high and the low USETECH groups demonstrated a statistically significant difference ($p = 1.079e\text{-}05$). While the t-test for HAPPY scores was not significant ($p = 0.3007$), the test for the other scores was. Finally, the Wilcoxon rank-sum test failed to show statistical significance in HAPPY scores between the groups ($p = 0.3196$).

The effect size of the significant difference between EDUC scores was computed using d Cohen. The effect size was 0.2004482, the Cohen's criteria defining this as a small to medium effect.

To find out the differences in mean HAPPY and EDUC scores, for RACE and MARITAL status, respectively, the one-way ANOVA tests were done. The ANOVA assumes normality and homoscedasticity which was tested using the Shapiro-Wilk and Levene's tests respectively. As the normality hypothesis was not followed, it was necessary to use the Kruskal-Wallis test as an alternative, which is a non-parametric test.

ANOVA and Kruskal-Wallis tests, thus, made it evident that there were statistically significant variations both in HAPPY and EDUC scores for various categories of RACE and MARITAL status (p-values < 0.05 for all). Tukey's HSD was applied post-hoc as a follow-up to the significant results of the ANOVA to differentiate between the specific groups.

Coefficient of determination (eta-squared) was used to calculate the effect size for the significant ANOVA results. Concerning the HAPPY scores, the effect sizes were 0.004 (RACE) and 0.081 (MARITAL), that were of small to medium magnitudes. In terms of EDUC, the effect sizes were 0.021 (RACE) and 0.018 (MARITAL), which reveal that these are both small effects.

Final Model

Multiple Linear Regression Model

A multiple linear regression model was fitted to examine the association between the variables HAPPY (dependent variable) and USETECH, AGE, SEX, and RACE (independent variables). The model was numbered as USETECH for technology use, AGE to consider differences in age, SEX for gender effects, and RACE to understand the racial differences.

The results of the multiple linear regression model revealed that only the RACE variable had a positive and significant effect on the HAPPY scores. Particularly, the Black race (to the White race) had higher HAPPY scores ($\beta = 0.1127718$, $p = 0.00224$). The other independent variables, namely, USETECH, AGE, and SEX, did not contribute to the HAPPY scores' significant values (all p-values greater than 0.05). The model as a whole was statistically significant ($F(5, 2335)= 3.02$, $p = 0.01011$), and the R-squared adjusted value of 0.004297 shows that it explains only a small proportion of the variance of HAPPY scores.

Binary Logistic Regression Model

A binary logistic regression model was fit to predict life satisfaction (LifeSat), a binary variable derived from the HAPPY variable, using the same set of independent variables: USE, TECH, AGES, SEXES, and RACES. The logistic model offers a possibility to interpret the influences of the independent variables on the probability of being "Very Happy" (LifeSat = 1) through odds ratios.

The logistic regression model results unveiled that AGE was the only statistically significant factor affecting life satisfaction. The odds ratio for AGE was 1.005371 ($p = 0.0319$), implying that older participants had a slightly higher probability of being "Very Happy" compared to their younger counterparts. However, the other independent variables, namely, USETECH, SEX, and RACE, were insignificant with respect to life satisfaction (all p-values > 0.0 The well-known model diagnostics statistics, such as the AIC (Akaike Information Criterion) and the count R-squared can be used to check the general model performance.

Results

The objectives of the analysis were designed to determine whether the application of technology (USETECH) among adolescents and children affects their academic performance (measured by educational attainment, EDUC) and life satisfaction (HAPPY). The data from the independent samples test showed a statistically significant difference ($p = 1.079e-05$, $d = 0.2004482$) in the EDUC score between the high and low USETECH groups, with a small medium effect size. People participating in the high USETECH group had a higher mean education level as compared to those in the low USETECH group.

But the inference demonstrates that no significant difference exists in the HAPPY scores between the high USETECH and the low USETECH groups, which is shown by the t-test ($p = 0.3007$) and the non-parametric Wilcoxon rank-sum test ($p = 0.3196$). This could mean that the influence of tech on life satisfaction of these participants does not have a direct relationship.

The multiple linear regression model that included USETECH, AGE, SEX, and RACE as independent variables further showed that the USETECH could not predict HAPPY scores with high level of significance ($p = 0.36193$). However, the model displayed the fact that black people (compared to white) had higher HAPPY scores ($p = 0.00224$).

In relation to life satisfaction, binary logistic regression model showed that AGE was the only significant predictor with older population having higher odds to be "Very Happy" (odds ratio = 1.005371, $p = 0.0319$). The technology (USETECH) did not significantly influence the probability of belonging among this group to the category "Very Happy" in the sample.

Discussions

The research question was focused on examining the relationship between the technology use by youths and its consequences on academics and life satisfaction among adolescents and children. These findings showed a complex pattern of interrelatedness among these variables. Despite the fact that higher tech use (USETECH) was significantly related to higher education (EDUC), it did not seem to have a direct effect on life satisfaction (HAPPY) in this specific sample. The positive correlation between the use of technology and academic performance conforms to the existing literature that rightly suggests that technology if used appropriately can be very instrumental in improving learning outcomes and educational attainment[2]. But, on the other hand, the absence of straight causal connection between technology use and contentment does not match some of the previous studies, which linked with the excessive technology usage and the lower well-being and happiness[3]. The fact that the results of the studies about technology use, academic performance, and subjective well-being are different shows the complicated connection between them.

It is necessary to note that, although the study has several strengths, they also come with some drawbacks. It could be that the sample is not representative of the entire population of the adolescents and children since it came from only specific survey. Furthermore, omitting the measure of technology use (USETECH) could leave out the width of the technology usage spectrum considering the types of activities performed and the time spent on different platforms.

Further studies are required to examine the different aspects of technology use beyond the quantitative indicators that are commonly used. These could include qualitative assessments of how technology is used for education and entertainment. Additionally, the longitudinal studies may offer the researchers many insights on whether the technology use affects student's academic performance and well-being at different stages of development.

Conclusion

In conclusion, it was concluded that the study was conducted on the use of technology and the influence it has on the academic performance and life quality of adolescents and children. The results suggest that technology usage tends to be prevalent among people with higher education, but it is not necessarily associated with the increased life satisfaction in this particular sample. While some variables like race and age showed more marked impacts on the level of happiness, others had far less impact.

These discoveries bring to the conversation of the influence of technology in learning and wellbeing. Whilst technology could be a useful instrument in boosting students' academic performance, the influence it has on the subjective well-being of the individual might be more intricate and vary from person to person and from one context to another.

It is necessary for future research to find better indicators of technology use and to analyze particular applications of technology in educational and private contexts. Besides, longitudinal studies might have the capability of adding more knowledge to the long-term effects of technology use on academic performance and psychological well-being at various developmental stages.

References

1. Lepp A, Barkley JE, Karpinski AC. The relationship between cell phone use, academic performance, anxiety, and Satisfaction with Life in college students. Comput Human Behav. 2014;(pagination):N/A. doi:10.1016/j.chb.2013.10.049
2. Moksnes, U.K., Løhre, A., Lillefjell, M. et al. The Association Between School Stress, Life Satisfaction and Depressive Symptoms in Adolescents: Life Satisfaction as a Potential Mediator. Soc Indic Res 125, 339–357 (2016). https://doi.org/10.1007/s11205-014-0842-0
3. Crede J, Wirthwein L, McElvany N and Steinmayr R (2015) Adolescents' academic achievement and life satisfaction: the role of parents' education. Front. Psychol. 6:52. doi: 10.3389/fpsyg.2015.00052
4. Żerebecki BG, Opree SJ. The direct and indirect effects of social technology use on children's life satisfaction. Int J Child Comput Interact. 2022;(pagination):N/A. doi:10.1016/j.ijcci.2022.100538
5. Sampasa-Kanyinga H, Hamilton HA, Goldfield GS, Chaput J-P. Problem Technology Use, Academic Performance, and School Connectedness among Adolescents. International Journal of Environmental Research and Public Health. 2022; 19(4):2337. https://doi.org/10.3390/ijerph19042337

Appendix

Final

2024-04-28

Load the required packages using library()

```
library(package = tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.
0.0 ──

## ✔ dplyr     1.1.4     ✔ readr     2.1.5

## ✔ forcats   1.0.0     ✔ stringr   1.5.1

## ✔ ggplot2   3.5.0     ✔ tibble    3.2.1

## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1

## ✔ purrr     1.0.2

## ── Conflicts ────────────────────────────────────────── tidyverse_conflict
s() ──

## ✖ dplyr::filter() masks stats::filter()

## ✖ dplyr::lag()    masks stats::lag()

## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(package = dplyr)

library(package = ggplot2)

library(package = haven)
```

Load the Data File

```
load("gss2018.rda")
```

Summary of the data file

```
summary(object = GSS)
```

```
##       YEAR          BALLOT         USETECH          HAPPY

##  Min.   :2018   Min.   :1.000   Min.   : -1.00   Min.   :1.000

##  1st Qu.:2018   1st Qu.:1.000   1st Qu.: -1.00   1st Qu.:1.000

##  Median :2018   Median :2.000   Median : 10.00   Median :2.000

##  Mean   :2018   Mean   :2.002   Mean   : 48.09   Mean   :1.855
```

```
##  3rd Qu.:2018   3rd Qu.:3.000   3rd Qu.: 80.00   3rd Qu.:2.000
##  Max.   :2018   Max.   :3.000   Max.   :999.00   Max.   :8.000
##     PARTYID         RINCOME            RACE             SEX
##  Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.: 0.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :3.000   Median : 9.000   Median :1.000   Median :2.000
##  Mean   :2.968   Mean   : 7.509   Mean   :1.394   Mean   :1.552
##  3rd Qu.:5.000   3rd Qu.:12.000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :9.000   Max.   :98.000   Max.   :3.000   Max.   :2.000
##      DEGREE           EDUC            AGE           MARITAL
##  Min.   :0.000   Min.   : 0.00   Min.   :18.00   Min.   :1.00
##  1st Qu.:1.000   1st Qu.:12.00   1st Qu.:34.00   1st Qu.:1.00
##  Median :1.000   Median :14.00   Median :48.00   Median :2.00
##  Mean   :1.684   Mean   :13.84   Mean   :49.13   Mean   :2.67
##  3rd Qu.:3.000   3rd Qu.:16.00   3rd Qu.:63.00   3rd Qu.:5.00
##  Max.   :4.000   Max.   :99.00   Max.   :99.00   Max.   :9.00
##       HRS2              HRS1          WRKSTAT          ID_
##  Min.   :-1.00000   Min.   :-1.00   Min.   :1.000   Min.   :   1
##  1st Qu.:-1.00000   1st Qu.:-1.00   1st Qu.:1.000   1st Qu.: 588
##  Median :-1.00000   Median :30.00   Median :2.000   Median :1176
##  Mean   : 0.08017   Mean   :24.47   Mean   :2.963   Mean   :1175
##  3rd Qu.:-1.00000   3rd Qu.:40.00   3rd Qu.:5.000   3rd Qu.:1762
##  Max.   :99.00000   Max.   :99.00   Max.   :9.000   Max.   :2348
##     UNHAPPY
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean   :1.039
##  3rd Qu.:2.000
##  Max.   :9.000
```

Selecting Variables of Interest and then recoding them

```
GSS_Cleaned <- GSS %>%
  select(EDUC, HAPPY, USETECH, AGE, SEX, RACE, WRKSTAT, RINCOME, MARITAL, HRS
1, HRS2) %>%
```

```
  mutate(USETECH = na_if(x = USETECH, y = -1)) %>%
  mutate(USETECH = na_if(x = USETECH, y = 999)) %>%
  mutate(USETECH = na_if(x = USETECH, y = 998)) %>%
  mutate(
    EDUC = factor(EDUC, levels = 0:20, labels = c(
      "No formal schooling", "1st grade", "2nd grade", "3rd grade", "4th grad
e",
      "5th grade", "6th grade", "7th grade", "8th grade", "9th grade", "10th
grade",
      "11th grade", "12th grade", "1 year of college", "2 years of college",
      "3 years of college", "4 years of college", "5 years of college",
      "6 years of college", "7 years of college", "8 years of college"
    )),
    HAPPY = factor(HAPPY, levels = 1:3, labels = c("Very Happy", "Pretty happ
y", "Not Too happy")),
    WRKSTAT = factor(WRKSTAT, levels = 1:9, labels = c(
      "Working Full Time", "Working Part Time", "With a job not working", "Un
employed",
      "Retired", "In School", "Keeping House", "Other", "No Answer"
    )),
    RINCOME = factor(RINCOME, levels = 1:13, labels = c(
      "Under $1000", "$1000 - $2999", "$3000 - $3999", "$4000 - $4999", "$500
0 - $5999",
      "$6000 - $6999", "$7000 - $7999", "$8000 - $9999", "$10000 - $14999", "
$15000 - $19000",
      "$20000 - $24999", "$25000 and over", "Refused"
    )),
    MARITAL = factor(MARITAL, levels = 1:5, labels = c(
      "Married", "Widowed", "Divorced", "Separated", "Never married"
    )),
    SEX = factor(SEX, levels = 1:2, labels = c("Male", "Female")),
    RACE = factor(RACE, levels = 1:3, labels = c("White", "Black", "Others"))
,
    HRS1 = factor(HRS1, levels = 0:10, labels = c(
      "0-9 hours", "10-19 hours", "20-29 hours", "30-39 hours", "40-49 hours"
,
      "50-59 hours", "60-69 hours", "70-79 hours", "80 or more hours", "Don't
know", "No answer"
```

```
    )),
    HRS2 = factor(HRS2, levels = 0:10, labels = c(
      "0-9 hours", "10-19 hours", "20-29 hours", "30-39 hours", "40-49 hours"
,
      "50-59 hours", "60-69 hours", "70-79 hours", "80 or more hours", "Don't
know", "No answer"
    ))
)
summary(GSS_Cleaned)
```

```
##                       EDUC                   HAPPY          USETECH              AGE
##   12th grade         :656   Very Happy   : 701   Min.   :  0.00   Min.   :18
.00
##   4 years of college:430   Pretty happy :1304   1st Qu.: 15.00   1st Qu.:34
.00
##   2 years of college:312   Not Too happy: 336   Median : 60.00   Median :48
.00
##   1 year of college :183   NA's         :   4   Mean   : 55.15   Mean   :49
.13
##   3 years of college:126                        3rd Qu.: 90.00   3rd Qu.:63
.00
##   (Other)            :635                        Max.   :100.00   Max.   :99
.00
##   NA's               :  3                        NA's   :936
##       SEX            RACE                     WRKSTAT                  RINCOME
##   Male   :1051   White :1692   Working Full Time:1133   $25000 and over:849
##   Female:1294   Black : 383   Retired          : 445   Refused         :108
##                  Others: 270   Working Part Time: 258   $20000 - $24999:107
##                                Keeping House    : 242   $10000 - $14999: 93
##                                Unemployed       :  84   $15000 - $19000: 61
##                                In School        :  81   (Other)         :202
##                                (Other)          : 102   NA's            :925
##             MARITAL                   HRS1                HRS2
##   Married      :998   No answer       :   9   60-69 hours:   1
##   Widowed      :200   60-69 hours     :   8   Don't know :   1
##   Divorced     :402   10-19 hours     :   7   No answer  :   1
##   Separated    : 75   80 or more hours:   6   0-9 hours  :   0
##   Never married:668   30-39 hours     :   5   10-19 hours:   0
##   NA's         :  2   (Other)         :   6   (Other)    :   0
```

```
##                        NA's            :2304   NA's        :2342
```

```
GSS_Cleaned$USETECH[is.na(GSS_Cleaned$USETECH)] <- mean(GSS_Cleaned$USETECH,
na.rm = TRUE)
```

Descriptive Statistics

```
table_one <- tableone::CreateTableOne(vars = c("EDUC", "HAPPY", "USETECH", "A
GE", "SEX", "RACE", "WRKSTAT", "RINCOME", "MARITAL", "HRS1", "HRS2"),
                            data = GSS_Cleaned,
                            factorVars = c("EDUC", "HAPPY", "SEX", "RACE", "W
RKSTAT", "RINCOME", "MARITAL", "HRS1", "HRS2"))


# Print the table
print(table_one, showAllLevels = TRUE)
```

```
##
##                         level                   Overall
##    n                                             2345
##    EDUC (%)             No formal schooling         4 ( 0.2)
##                         1st grade                   2 ( 0.1)
##                         2nd grade                   4 ( 0.2)
##                         3rd grade                  10 ( 0.4)
##                         4th grade                   5 ( 0.2)
##                         5th grade                   3 ( 0.1)
##                         6th grade                  20 ( 0.9)
##                         7th grade                   8 ( 0.3)
##                         8th grade                  35 ( 1.5)
##                         9th grade                  51 ( 2.2)
##                         10th grade                 65 ( 2.8)
##                         11th grade                 95 ( 4.1)
##                         12th grade                656 (28.0)
##                         1 year of college         183 ( 7.8)
##                         2 years of college        312 (13.3)
##                         3 years of college        126 ( 5.4)
##                         4 years of college        430 (18.4)
##                         5 years of college         97 ( 4.1)
##                         6 years of college        119 ( 5.1)
##                         7 years of college         45 ( 1.9)
```

```
##                        8 years of college        72 ( 3.1)
##   HAPPY (%)            Very Happy               701 (29.9)
##                        Pretty happy            1304 (55.7)
##                        Not Too happy            336 (14.4)
##   USETECH (mean (SD))                         55.15 (29.32)
##   AGE (mean (SD))                             49.13 (18.24)
##   SEX (%)              Male                    1051 (44.8)
##                        Female                  1294 (55.2)
##   RACE (%)             White                   1692 (72.2)
##                        Black                    383 (16.3)
##                        Others                   270 (11.5)
##   WRKSTAT (%)          Working Full Time       1133 (48.3)
##                        Working Part Time        258 (11.0)
##                        With a job not working    52 ( 2.2)
##                        Unemployed                84 ( 3.6)
##                        Retired                  445 (19.0)
##                        In School                 81 ( 3.5)
##                        Keeping House            242 (10.3)
##                        Other                     48 ( 2.0)
##                        No Answer                  2 ( 0.1)
##   RINCOME (%)          Under $1000               33 ( 2.3)
##                        $1000 - $2999             32 ( 2.3)
##                        $3000 - $3999             32 ( 2.3)
##                        $4000 - $4999             21 ( 1.5)
##                        $5000 - $5999             21 ( 1.5)
##                        $6000 - $6999             12 ( 0.8)
##                        $7000 - $7999             18 ( 1.3)
##                        $8000 - $9999             33 ( 2.3)
##                        $10000 - $14999           93 ( 6.5)
##                        $15000 - $19000           61 ( 4.3)
##                        $20000 - $24999          107 ( 7.5)
##                        $25000 and over          849 (59.8)
##                        Refused                  108 ( 7.6)
##   MARITAL (%)          Married                  998 (42.6)
```

```
##                        Widowed                 200 ( 8.5)
##                        Divorced                402 (17.2)
##                        Separated                75 ( 3.2)
##                        Never married           668 (28.5)
##     HRS1 (%)           10-19 hours               7 (17.1)
##                        20-29 hours               2 ( 4.9)
##                        30-39 hours               5 (12.2)
##                        50-59 hours               2 ( 4.9)
##                        60-69 hours               8 (19.5)
##                        70-79 hours               2 ( 4.9)
##                        80 or more hours          6 (14.6)
##                        No answer                 9 (22.0)
##     HRS2 (%)           60-69 hours               1 (33.3)
##                        Don't know                1 (33.3)
##                        No answer                 1 (33.3)
```

Bar Chart for Education

```
ggplot(GSS_Cleaned, aes(x = EDUC)) +
  geom_bar() +
  labs(title = "Distribution of Education Level", x = "Education Level", y =
"Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip() +
  theme(plot.title = element_text(hjust = 0.5))
```

Bar Chart for Happiness Level

```
ggplot(GSS_Cleaned, aes(x = HAPPY)) +
  geom_bar() +
  labs(title = "Distribution of Happiness Level", x = "Happiness Level", y =
"Count")
```

Scatterplot with regression line for USETECH and Happy

```
ggplot(GSS_Cleaned, aes(x = USETECH, y = as.numeric(HAPPY))) +
```

```
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relationship between USETECH and Happiness", x = "USETECH", y
= "Happiness")
```

```
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 4 rows containing missing values or values outside the sc
ale range
## (`geom_point()`).
```

Scatterplot for Usetech and Education

```
ggplot(GSS_Cleaned, aes(x = USETECH, y = as.numeric(EDUC))) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relationship between USETECH and Education", x = "USETECH", y
= "Education")
```

```
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 3 rows containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 3 rows containing missing values or values outside the sc
ale range
## (`geom_point()`).
```

Histogram for Age

```
ggplot(GSS_Cleaned, aes(x = AGE)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "white") +
  labs(title = "Distribution of Age", x = "Age", y = "Count")
```

Statistical Tests

Independent samples t-tests

```
# Create high and low USETECH groups
```

```
GSS_Cleaned$USETECH_group <- ifelse(GSS_Cleaned$USETECH >= median(GSS_Cleaned
$USETECH, na.rm = TRUE), "High", "Low")
```

Null Hypothesis: There is no significant difference in mean HAPPY between high and low USETECH groups. Alternate Hypothesis: There is a significant difference in mean HAPPY between high and low USETECH groups.

```
# T-test for HAPPY
t_test_happy <- t.test(as.numeric(HAPPY) ~ USETECH_group, data = GSS_Cleaned)
print(t_test_happy)
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(HAPPY) by USETECH_group
## t = 1.0355, df = 1268, p-value = 0.3007
## alternative hypothesis: true difference in means between group High and gr
oup Low is not equal to 0
## 95 percent confidence interval:
##  -0.02716859  0.08790388
## sample estimates:
## mean in group High  mean in group Low
##           1.852853           1.822485
```

Here, we retain the null hypothesis and conclude that there is no significant difference in mean HAPPY between high and low USETECH groups.

Null Hypothesis: There is no significant difference in EDUC scores between high and low USETECH groups. Alternate Hypothesis: There is a significant difference in EDUC scores between high and low USETECH groups.

```
# T-test for EDUC
t_test_educ <- t.test(as.numeric(EDUC) ~ USETECH_group, data = GSS_Cleaned)
print(t_test_educ)
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(EDUC) by USETECH_group
## t = 4.4185, df = 1264.5, p-value = 1.079e-05
## alternative hypothesis: true difference in means between group High and gr
oup Low is not equal to 0
```

```
## 95 percent confidence interval:
##  0.3303631 0.8580164
## sample estimates:
## mean in group High  mean in group Low
##           14.90336           14.30917
```

Here we reject the null hypothesis and conclude that there is a significant difference in EDUC scores between high and low USETECH groups.

Assumption Homogenity of variances assumption

```
library(car)
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##     recode
## The following object is masked from 'package:purrr':
##
##     some
leveneTest(as.numeric(HAPPY) ~ USETECH_group, data = GSS_Cleaned)
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group    1  0.0177 0.8942
##       2339
leveneTest(as.numeric(EDUC) ~ USETECH_group, data = GSS_Cleaned)
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value  Pr(>F)
## group    1  3.1587 0.07565 .
##       2340
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assumption of homogenity

Assumption of Normality

```
shapiro.test(as.numeric(GSS_Cleaned$HAPPY[GSS_Cleaned$USETECH_group == "High"
]))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  as.numeric(GSS_Cleaned$HAPPY[GSS_Cleaned$USETECH_group == "High"])
## W = 0.78779, p-value < 2.2e-16
```

```
shapiro.test(as.numeric(GSS_Cleaned$HAPPY[GSS_Cleaned$USETECH_group == "Low"]
))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  as.numeric(GSS_Cleaned$HAPPY[GSS_Cleaned$USETECH_group == "Low"])
## W = 0.78189, p-value < 2.2e-16
```

```
shapiro.test(as.numeric(GSS_Cleaned$EDUC[GSS_Cleaned$USETECH_group == "High"]
))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  as.numeric(GSS_Cleaned$EDUC[GSS_Cleaned$USETECH_group == "High"])
## W = 0.94508, p-value < 2.2e-16
```

```
shapiro.test(as.numeric(GSS_Cleaned$EDUC[GSS_Cleaned$USETECH_group == "Low"])
)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  as.numeric(GSS_Cleaned$EDUC[GSS_Cleaned$USETECH_group == "Low"])
## W = 0.93654, p-value = 2.322e-16
```

As Assumption of normality is violated

```
wilcox.test(as.numeric(HAPPY) ~ USETECH_group, data = GSS_Cleaned)
```

```
## 
```

```
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(HAPPY) by USETECH_group
## W = 575943, p-value = 0.3196
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(as.numeric(EDUC) ~ USETECH_group, data = GSS_Cleaned)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(EDUC) by USETECH_group
## W = 628997, p-value = 6.333e-06
## alternative hypothesis: true location shift is not equal to 0
```

Anova Test Null Hypothesis: There are no significant differences in mean HAPPY across RACE subgroups. Alternate Hypothesis: There are significant differences in mean HAPPY across RACE and subgroups.

```
# ANOVA for HAPPY across RACE
anova_happy_race <- aov(as.numeric(HAPPY) ~ RACE, data = GSS_Cleaned)
summary(anova_happy_race)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## RACE           2    4.5  2.2412   5.371 0.00471 **
## Residuals   2338  975.6  0.4173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 4 observations deleted due to missingness
```

Here, we reject the null hypothesis and conclude that there are significant differences in mean HAPPY across RACE and subgroups.

Null Hypothesis: There are no significant differences in mean HAPPY across MARITAL subgroups. Alternate Hypothesis: There are significant differences in mean HAPPY across MARITAL subgroups.

```
# ANOVA for HAPPY across MARITAL
anova_happy_marital <- aov(as.numeric(HAPPY) ~ MARITAL, data = GSS_Cleaned)
summary(anova_happy_marital)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## MARITAL        4   83.1  20.763   54.07 <2e-16 ***
```

```
## Residuals   2334  896.3   0.384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 6 observations deleted due to missingness
```

Here, we reject the null hypothesis and conclude There are significant differences in mean HAPPY across MARITAL subgroups.

Null Hypothesis: There are no significant differences EDUC scores across RACE subgroups.
Alternate Hypothesis: There are significant differences in EDUC scores across RACE subgroups.

```
# ANOVA for EDUC across RACE
anova_educ_race <- aov(as.numeric(EDUC) ~ RACE, data = GSS_Cleaned)
summary(anova_educ_race)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## RACE           2    451  225.69   26.03 6.6e-12 ***
## Residuals   2339  20280    8.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3 observations deleted due to missingness
```

Here, we reject the null hypothesis and conclude, There are significant differences in EDUC scores across RACE subgroups.

Null Hypothesis: There are no significant differences EDUC scores across MARITAL subgroups.
Alternate Hypothesis: There are significant differences EDUC scores across MARITAL subgroups.

```
# ANOVA for EDUC across MARITAL
anova_educ_marital <- aov(as.numeric(EDUC) ~ MARITAL, data = GSS_Cleaned)
summary(anova_educ_marital)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## MARITAL        4    375   93.74   10.77 1.19e-08 ***
## Residuals   2335  20326    8.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 5 observations deleted due to missingness
```

Here, we reject the null hypothesis and conclude, There are significant differences EDUC scores across MARITAL subgroups.

As normality assumption is violated

```
kruskal.test(as.numeric(HAPPY) ~ RACE, data = GSS_Cleaned)
##
##  Kruskal-Wallis rank sum test
##
## data:  as.numeric(HAPPY) by RACE
## Kruskal-Wallis chi-squared = 9.2336, df = 2, p-value = 0.009884
kruskal.test(as.numeric(HAPPY) ~ MARITAL, data = GSS_Cleaned)
##
##  Kruskal-Wallis rank sum test
##
## data:  as.numeric(HAPPY) by MARITAL
## Kruskal-Wallis chi-squared = 198.55, df = 4, p-value < 2.2e-16
kruskal.test(as.numeric(EDUC) ~ RACE, data = GSS_Cleaned)
##
##  Kruskal-Wallis rank sum test
##
## data:  as.numeric(EDUC) by RACE
## Kruskal-Wallis chi-squared = 31.423, df = 2, p-value = 1.502e-07
kruskal.test(as.numeric(EDUC) ~ MARITAL, data = GSS_Cleaned)
##
##  Kruskal-Wallis rank sum test
##
## data:  as.numeric(EDUC) by MARITAL
## Kruskal-Wallis chi-squared = 42.908, df = 4, p-value = 1.081e-08
```

As all the tests are significant, indicating differences in the means of HAPPY and EDUC across the levels of RACE and MARITAL.

We are conducting Tukey's HSD post-hoc tests

```
TukeyHSD(anova_happy_race)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = as.numeric(HAPPY) ~ RACE, data = GSS_Cleaned)
##
```

```
## $RACE
##                      diff         lwr        upr     p adj
## Black-White   0.11924897  0.03324445 0.20525348 0.0033342
## Others-White  0.03903454 -0.06024961 0.13831869 0.6263871
## Others-Black -0.08021442 -0.20079512 0.04036627 0.2633041
```

```
TukeyHSD(anova_happy_marital)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = as.numeric(HAPPY) ~ MARITAL, data = GSS_Cleaned)
##
## $MARITAL
##                                  diff         lwr       upr     p adj
## Widowed-Married          0.2971163490  0.16604463 0.4281881 0.0000000
## Divorced-Married         0.3895292844  0.28958160 0.4894770 0.0000000
## Separated-Married        0.4261704031  0.22234619 0.6299946 0.0000001
## Never married-Married    0.3886328656  0.30397137 0.4732944 0.0000000
## Divorced-Widowed         0.0924129353 -0.05397143 0.2387973 0.4196461
## Separated-Widowed        0.1290540541 -0.10112665 0.3592348 0.5425821
## Never married-Widowed    0.0915165165 -0.04488881 0.2279218 0.3555186
## Separated-Divorced       0.0366411187 -0.17735164 0.2506339 0.9902180
## Never married-Divorced  -0.0008964188 -0.10774284 0.1059500 0.9999999
## Never married-Separated -0.0375375375 -0.24483184 0.1697568 0.9879080
```

```
TukeyHSD(anova_educ_race)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = as.numeric(EDUC) ~ RACE, data = GSS_Cleaned)
##
## $RACE
##                     diff       lwr        upr     p adj
## Black-White  -0.6145841 -1.005782 -0.2233861 0.0006858
## Others-White -1.2970614 -1.750356 -0.8437667 0.0000000
## Others-Black -0.6824773 -1.232123 -0.1328311 0.0101372
```

```
TukeyHSD(anova_educ_marital)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = as.numeric(EDUC) ~ MARITAL, data = GSS_Cleaned)
##
## $MARITAL
##                                diff         lwr         upr     p adj
## Widowed-Married          -1.0384422 -1.66258906 -0.41429536 0.0000574
## Divorced-Married         -0.3944124 -0.87040803  0.08158331 0.1577414
## Separated-Married        -1.7151089 -2.67955747 -0.75066028 0.0000127
## Never married-Married    -0.5135919 -0.91647192 -0.11071190 0.0046401
## Divorced-Widowed          0.6440299 -0.05291689  1.34097659 0.0859763
## Separated-Widowed        -0.6766667 -1.76722798  0.41389464 0.4379149
## Never married-Widowed     0.5248503 -0.12436090  1.17406150 0.1773666
## Separated-Divorced       -1.3206965 -2.33377922 -0.30761382 0.0034874
## Never married-Divorced   -0.1191796 -0.62759648  0.38923737 0.9685082
## Never married-Separated   1.2015170  0.22066174  2.18237220 0.0074813
```

Effect Sizes

```
library(lsr)
cohensD(as.numeric(HAPPY) ~ USETECH_group, data = GSS_Cleaned)
## [1] 0.04692361
cohensD(as.numeric(EDUC) ~ USETECH_group, data = GSS_Cleaned)
## [1] 0.2004482
```

Multiple Linear Regression

```
model_happy <- lm(as.numeric(HAPPY) ~ USETECH + AGE + SEX + RACE, data = GSS_
Cleaned)
summary(model_happy)
##
## Call:
## lm(formula = as.numeric(HAPPY) ~ USETECH + AGE + SEX + RACE,
##     data = GSS_Cleaned)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0022 -0.7954  0.1512  0.2003  1.2404
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.9200618  0.0507766  37.814  < 2e-16 ***
## USETECH     -0.0004186  0.0004590  -0.912  0.36193
## AGE         -0.0013351  0.0007395  -1.805  0.07116 .
## SEXFemale   -0.0157863  0.0269582  -0.586  0.55821
## RACEBlack    0.1127718  0.0368610   3.059  0.00224 **
## RACEOthers   0.0268583  0.0427607   0.628  0.52999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6458 on 2335 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.006425,   Adjusted R-squared:  0.004297
## F-statistic:  3.02 on 5 and 2335 DF,  p-value: 0.01011
```

```r
GSS_Cleaned$LifeSat <- ifelse(GSS_Cleaned$HAPPY == "Very Happy", 1, 0)

# Binary Logistic Regression for Life Satisfaction
model_lifesat <- glm(LifeSat ~ USETECH + AGE + SEX + RACE, data = GSS_Cleaned
, family = binomial(link = "logit"))
summary(model_lifesat)
```

```
##
## Call:
## glm(formula = LifeSat ~ USETECH + AGE + SEX + RACE, family = binomial(link
= "logit"),
##     data = GSS_Cleaned)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.1482260  0.1735666  -6.615  3.7e-11 ***
## USETECH      0.0004916  0.0015640   0.314   0.7533
```

```
## AGE          0.0053569  0.0024970   2.145   0.0319 *
## SEXFemale    0.0296148  0.0913091   0.324   0.7457
## RACEBlack   -0.1347356  0.1277011  -1.055   0.2914
## RACEOthers   0.0937277  0.1429816   0.656   0.5121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2857.9  on 2340  degrees of freedom
## Residual deviance: 2851.2  on 2335  degrees of freedom
##   (4 observations deleted due to missingness)
## AIC: 2863.2
##
## Number of Fisher Scoring iterations: 4
```

```
odds.n.ends::odds.n.ends(model_lifesat)
```

```
## Waiting for profiling to be done...
## $`Logistic regression model significance`
## Chi-squared        d.f.           p
##       6.662       5.000       0.247
##
## $`Contingency tables (model fit): frequency predicted`
##                   Number observed
## Number predicted    1    0  Sum
##               1     0    0    0
##               0   701 1640 2341
##               Sum 701 1640 2341
##
## $`Count R-squared (model fit): percent correctly predicted`
## [1] 70.05553
##
## $`Model sensitivity`
## [1] 0
##
```

```
## $`Model specificity`
## [1] 1
##
## $`Predictor odds ratios and 95% CI`
##                   OR     2.5 %    97.5 %
## (Intercept) 0.317199 0.2252001 0.4448096
## USETECH     1.000492 0.9974348 1.0035711
## AGE         1.005371 1.0004616 1.0103060
## SEXFemale   1.030058 0.8614462 1.2322894
## RACEBlack   0.873947 0.6782144 1.1193318
## RACEOthers  1.098261 0.8268794 1.4492046
```