Youth Technology Use & Behavioral Health
Daniel Scheer & Gaetano Hirshout
B104 Final Project

Table of Contents

Table of Figures

Abstract

Introduction
To inform the potential for a relationship between the use of digital technology and adverse behavioral health outcomes two large representative datasets of American Youths were examined. From each two questions were selected, one to establish ordinal categorical data for time spent interacting with digital technology, and one to represent behavioral health outcomes.

Purpose
The Youth Risk Behavior Survey (YRBS) administered by the Centers for Disease Control and Prevention has since 1991 collected data [1] biennially from a representative sample of American Youths. Its purpose is to monitor the effectiveness of public health interventions by observing health-related behaviors [2]. In 2021 the United States Surgeon General issued an urgent public health advisory stating that "from 2009 to 2019, the proportion of high school students reporting persistent feelings of sadness or hopelessness increased by 40%; the share seriously considering attempting suicide increased by 36%; and the share creating a suicide plan increased by 44%" [3]. While this trend of increasing negative indicators of behavioral health outcomes has been observed to coincide with an increase in the use of digital technology there is little convincing evidence to establish causation [4]. As the YRBS question regarding technology use is broad based, and therefore lacks specificity to the type of media that young people are interacting with, an additional large dataset was also evaluated to provide data specific to social media use. The Monitoring the Future (MTF) Public-Use Cross-Sectional Datasets [5] was examined with permission from the University of Michigan.

Research Question
Is there a difference in self-reported indicators of adverse behavioral health outcomes among self-reported heavy users of digital technology?

Hypothesis

YRBS Data
H0 Null Hypothesis
There is no relationship between persistently feeling sad or hopeless (**question 25**), and time spent using digital technology (**question 80**).

H1 Alternative Hypothesis
There is a relationship between persistently feeling sad or hopeless (**question 25**), and time spent using digital technology (**question 80**).

MTF Data
H0 Null Hypothesis
There is no relationship between happiness (**question V7302**), and time spent on social networking sites (**question V7685**).

H1 Alternative Hypothesis
There is a relationship between happiness (**question V7302**), and time spent on social networking sites (**question V7685**).

Methodology

Data Retrieval & Winnowing:

Raw data retrieved in Microsoft Access format, winnowed via SQL query, imported into python via pandas library script, and assigned to a python data frame (see figure 1 below).
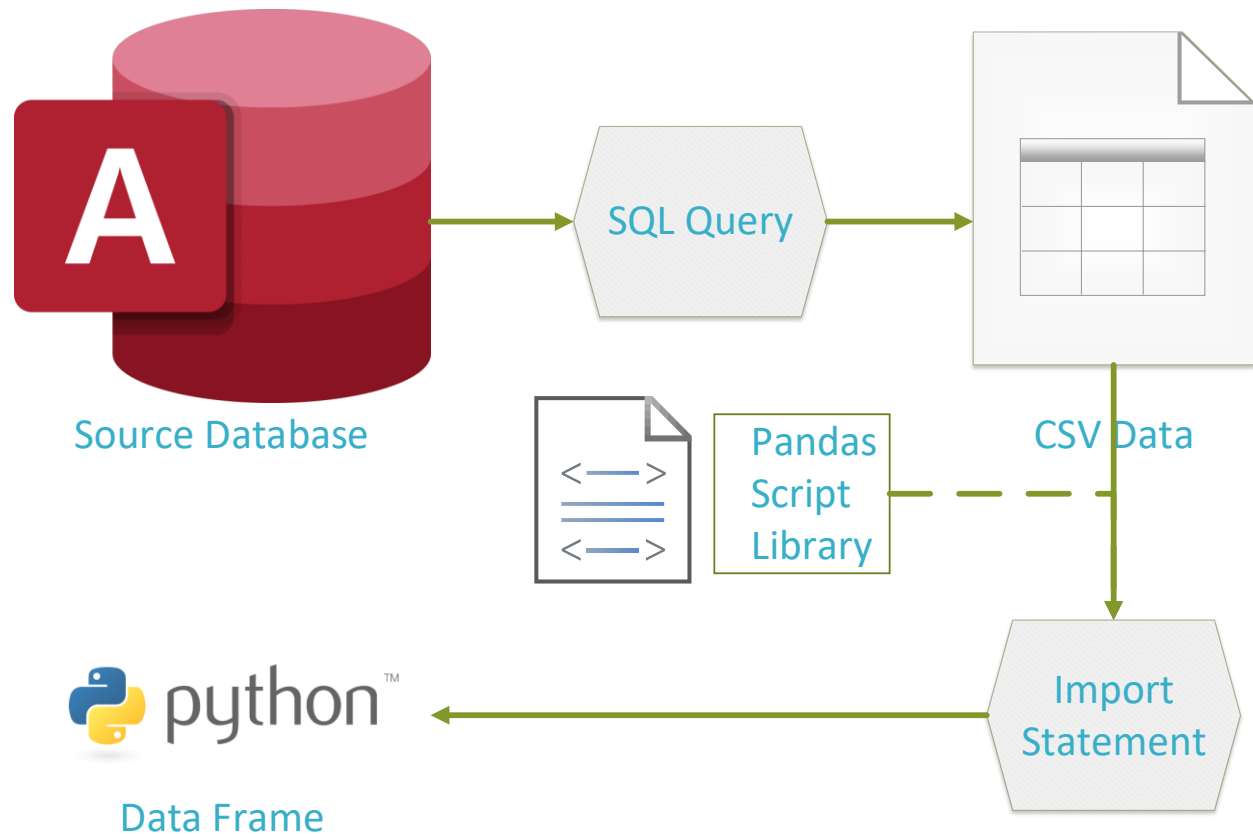
Figure 1 data retrieval & winnowing flowchart

Data Analysis:

Actual distribution among respondents was established using a stacked histogram. Expected distribution was calculated using chi-squared test of independence to determine if a relationship was present. A P-value of .05 was determined in advance to test the hypothesis.

Survey Questions

YRBS

Q80. On an average school day, how many hours do you play video or computer games or use a computer for something that is not school work? (Count time spent playing games, watching videos, texting, or using social media on your smartphone, computer, Xbox, PlayStation, iPad, or other tablet.)

A. I do not play video or computer games or use a computer for something that is not school work
B. Less than 1 hour per day

C. 1 hour per day
D. 2 hours per day
E. 3 hours per day
F. 4 hours per day
G. 5 or more hours per day

Variable label: Computer use

Q25. During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?

A. Yes
B. No

Variable label: Sad or hopeless

MTF

V7685: About how many hours on an average DAY do you spend on social networking sites like Facebook, Twitter, Instagram, etc.?

1="None"
2="Less than 1 hour"
3="1-2 hours"
4="3-4 hours"
5="5-6 hours"
6="7-8 hours"
7="9 hours or more"

V7302: Taking all things together, how would you say things are these days -- would you say you're very happy, pretty happy, or not too happy these days?

1="Not too happy"
2="Pretty happy"
3="Very happy"

Results
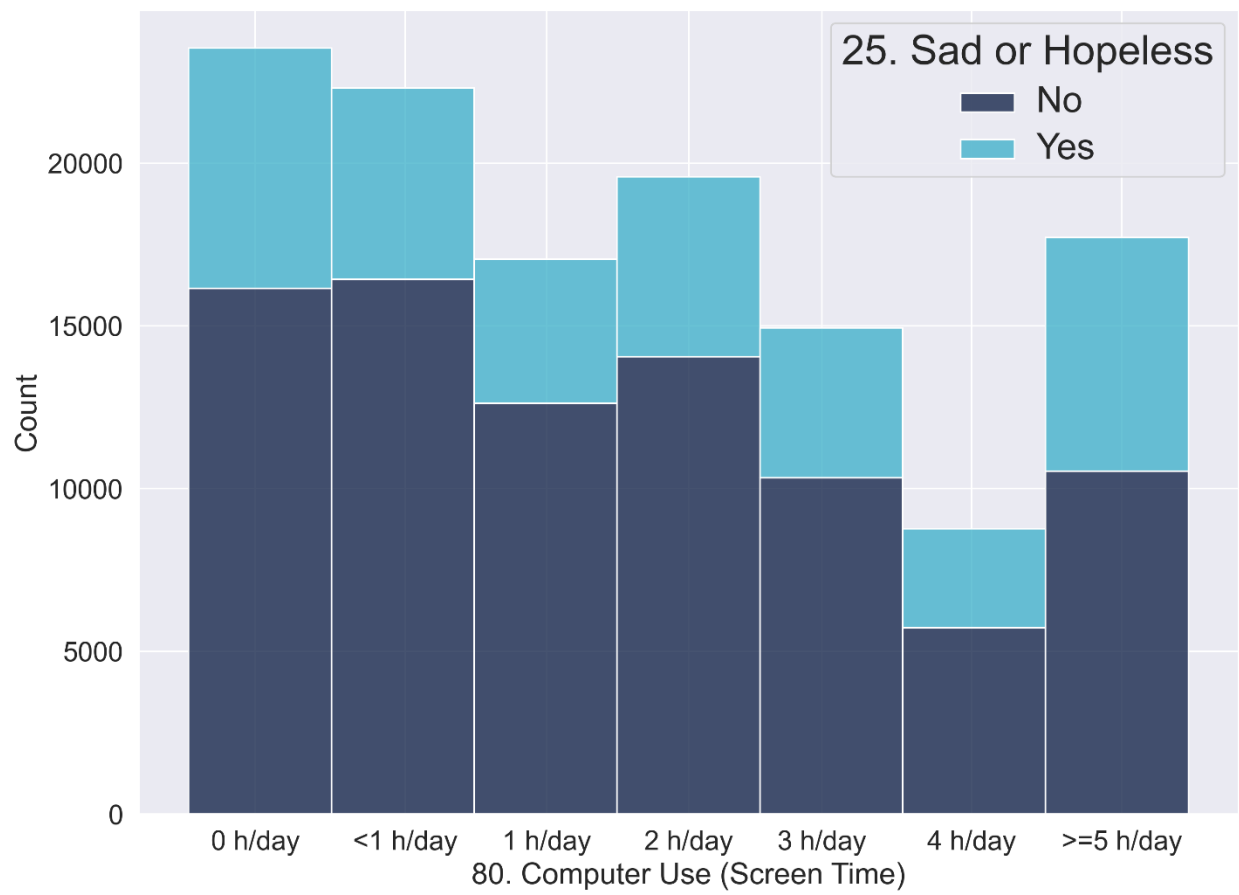
Figure 2 actual distribution stacked histogram for YRBS

Chi-Squared Test of Independence (YRBS)	
P-Value	Degrees of Freedom
4.2993388683043516e-280	6
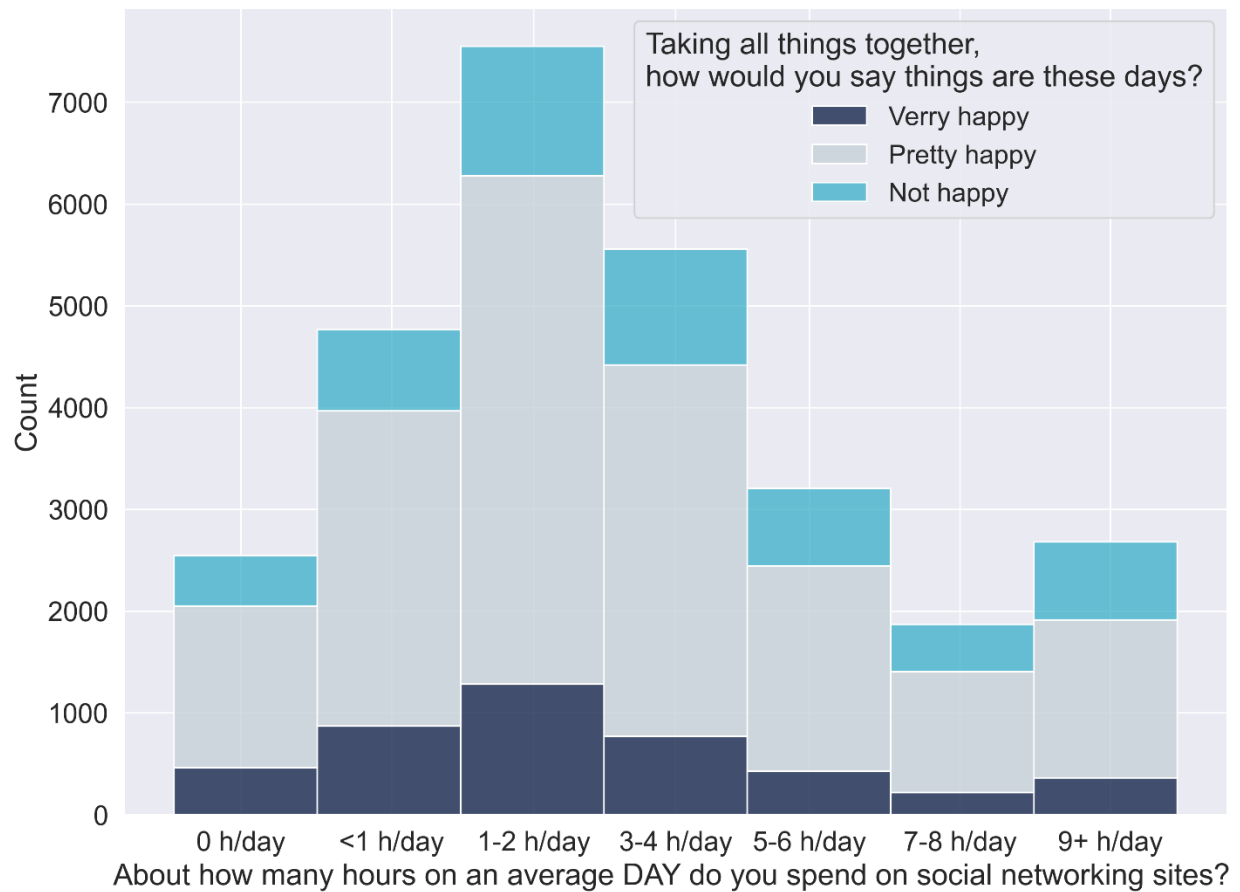
Figure 3 actual distribution stacked histogram for MTF

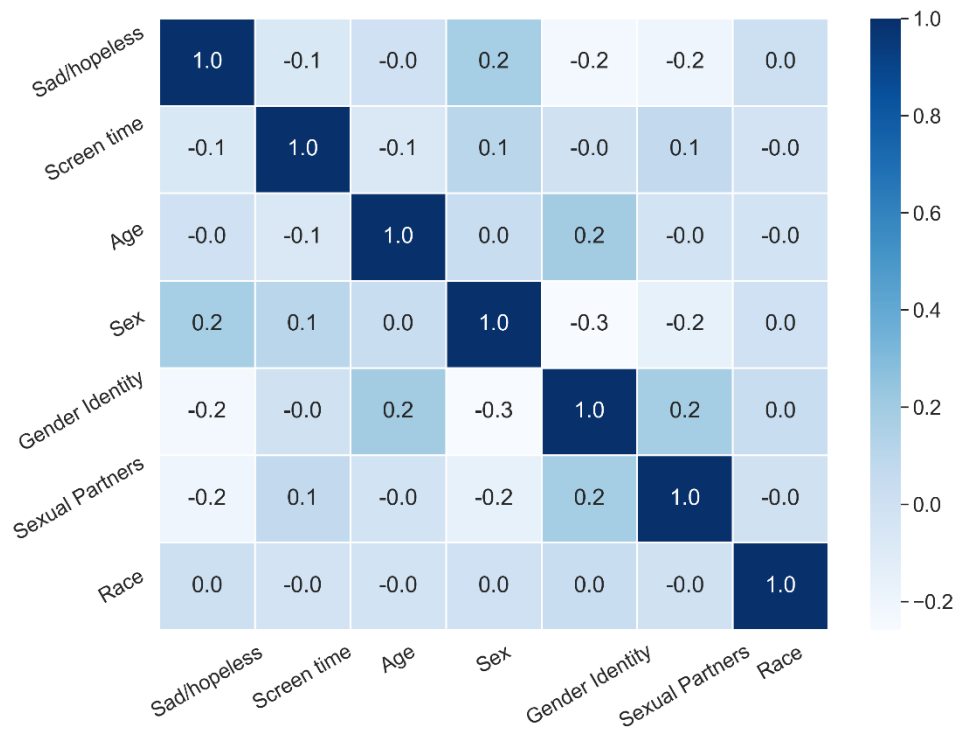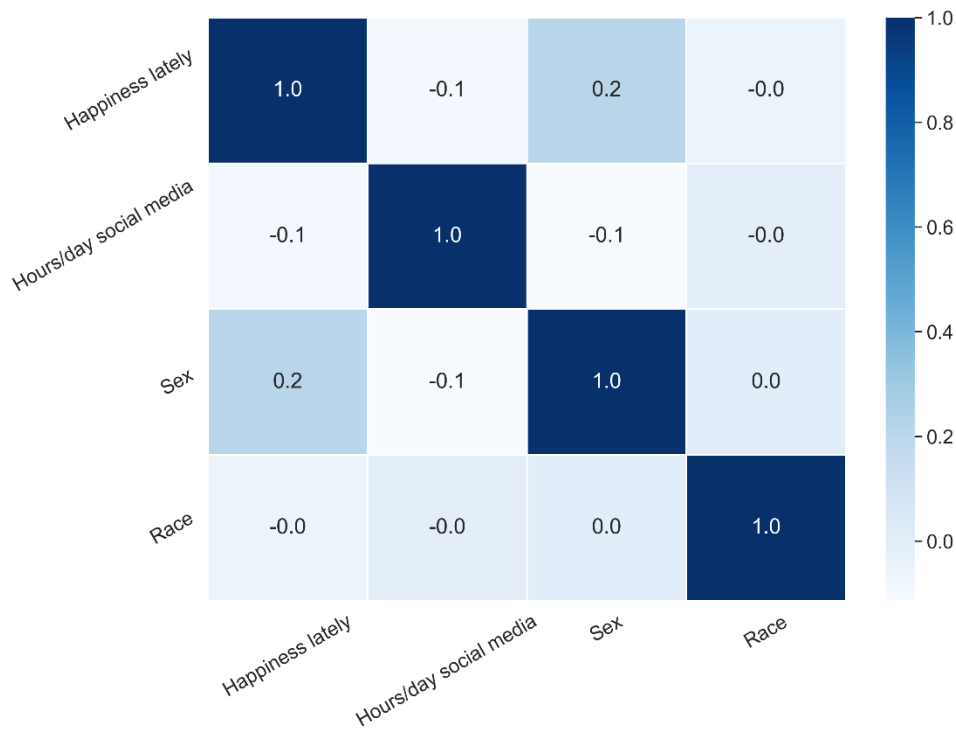| Chi-Squared Test of Independence (MTF)	
P-Value	Degrees of Freedom
2.8233542012215992e-61	12

Figure 4 heatmap for YRBS

Figure 5 heatmap for MTF

Limitations

The dataset captures screen time/social media use as ordinal categorical data, and self-reported indicators of behavioral health outcomes as categorical data. As such, the heatmaps generated in this this study via a correlation function are not sufficient to establish correlation due to the underlying data. They are provided for representative purposes to outline potential relationships in the data that may warrant further examination.

In the case of both datasets, the 2019 data was studied. The authors made a formal request to the Centers for Disease Control and Prevention for the 2021 (latest) YRBS dataset, however it will not be released until late April 2023 as the state level data was still being aggregated at the time of this writing.

Discussion/Implications

The use of digital technology has become so pervasive among young Americans that daily active users represent a significant portion of the sample size. The methods applied in this study found no significant relationships between screen time and behavioral health outcomes. These findings do not however negate the need for a harm reduction approach in designing technology for adolescents [6]. Nor could the underlying data's focus on screen time as a primary metric for digital media interaction fully account for the impacts of "dispositional, developmental, and social-context differences among media users" [7]. Considering the quantity of user data that is commoditized in the furtherance of profits by technology companies, the lack of data available to study digital harm among vulnerable populations is of great concern [8]. With 95% of teens reporting access to a smartphone [9], studying the social and behavioral impacts should be met with increased determination, better data collection, and transparency.

Python Code

```
##Final Project V2.1
#Class: CSCI/ISAT B104 - Spring
#Authors:
    #Scheer, Daniel
    #Hirshout, Gaetano

def main():

    #Import of assets
    import pandas as pd
    import seaborn as sns
    import matplotlib.pyplot as plt
    import scipy.stats as sp

    #Import of dataset's to pandas dataframe
    yrbs_dataset = pd.read_csv('Final_filter_YRBSS_dataset.csv', sep = ',')
    mtf_dataset = pd.read_csv('Final_filter_MTF_dataset.csv', sep = ',')

    #sns.set(font_scale = 1.5)

    #-------------------------------------------------- Settings --------------------------------------------------#
    #Menu Settings:
```

```python
#printing and indexing of graphical functions will be counted from left to right in menu loops:
show_functions = ['YRBSS - Heatmap - filtered questions', 'YRBSS - Stacked Histogram - Y:count X:Q80 Hue:Q25',
                'MTF - Heatmap - full datatset', 'MTF - Stacked Histogram - Y:count X:V7685 Hue:V7302',
                'YRBS - Chi square - result', 'MTF - Chi square - result']

#------------------------------------------------- YRBS Settings -------------------------------------------------#

#name list for questions in dataset: (Long form)
yrbs_question_names = ['25. Sad or Hopeless','80. Computer Use (Screen Time)']

#Column names for heatmap xy index
yrbs_col_names = ['Sad/hopeless', 'Screen time','Age', 'Sex' ,'Gender Identity', 'Sexual Partners', 'Race']

#Index for anwser values in quesiton 80 - YRBS dataset
q80_index = ['0 h/day', '<1 h/day', '1 h/day', '2 h/day', '3 h/day', '4 h/day', '>=5 h/day']

#Index for anwser values in question 25 - YRBS dataset
q25_index = ['No','Yes']

#------------------------------------------------- MTF Settings -------------------------------------------------#
#name list for questions in dataset: (Long form)
mtf_question_names = ['Taking all things together, \nhow would you say things are these days?',
'About how many hours on an average DAY do you spend on social networking sites? ']

#Column names for heatmap xy index
mtf_col_names = ['Happiness lately', 'Hours/day social media', 'Sex', 'Race']

#Index for anwser values in quesiton V7685 - MTF dataset
v7685_index = ['0 h/day', '<1 h/day', '1-2 h/day', '3-4 h/day', '5-6 h/day', '7-8 h/day', '9+ h/day']

#Index for anwser values in quesiton V7302 - MTF dataset
v7302_index = ['Verry happy', 'Pretty happy', 'Not happy']

#----------------------------------------------------------------------------------------------------------------#
#YRBSS dataset functions


#YRBS stacked histogram
def yrbs_stacked_histogram(col_names, q80_index, q25_index):

    #Set subplot size and res
    f, ax = plt.subplots(figsize = (12, 9), dpi = 300)

    #Color scheme for hues in histogram
    colors = {1:'#38AFC9', 2:'#081A40'}
```

```python
        #Plot histogram using y = count x = question 80, hue = question 25
        sns.histplot(data = yrbs_dataset, stat = "count", multiple = "stack", element = "bars", legend = True,
kde = False , bins = 7, palette = colors, hue = yrbs_dataset.q25, x = yrbs_dataset.q80)

        #Set question name x axis + spacing of xtick lables
        ax.set(xlabel = col_names[1], xticks = ([1.4, 2.3, 3.15, 4, 4.8, 5.7, 6.6]), xticklabels = q80_index)

        #set hue question name and anwser index
        ax.legend(title=col_names[0],labels = q25_index)

        plt.setp(ax.get_legend().get_texts(), fontsize='22')
        plt.setp(ax.get_legend().get_title(), fontsize='25')

        #Plot and show graph
        ax.plot()
        plt.show()

    #YRBS heatmap
    def yrbs_heatmap_full(yrbs_col_names):
        #Set figure size and res
        plt.figure(figsize=(12, 9), dpi = 300)

        #Filter and make dataset readable for heatmap
        yrbs_filter = yrbs_dataset[['q25', 'q80', 'age', 'sex', 'q65', 'q66', 'race7']].corr()

        #Create heatmap + set x&y lables
        sns.heatmap(data = yrbs_filter, fmt = '.1f', linewidths = 1, cmap='Blues', annot = True, xticklabels=
yrbs_col_names, yticklabels= yrbs_col_names)
        #Rotate names 30deg
        plt.xticks(rotation=30)
        plt.yticks(rotation=30)

        #show plot
        plt.show()


    #-----------------------------------------------------------------------------------------------------------#
    #MTF dataset functions:

    #MTF stacked histogram
    def mtf_stacked_histogram(col_names, v7685_index, v7302_index):
        #Set subplot size and resolution
        f, ax = plt.subplots(figsize = (12, 9), dpi = 300)

        #Set colors for hues in histogram
        colors = {1:'#38AFC9', 2:'#C4D0D7' ,3:'#081A40'}
```

```python
    #Create histogram
    sns.histplot(data = mtf_dataset, stat = "count", multiple = "stack", element = "bars", legend = True,
kde = False, bins = 7, palette = colors, hue = mtf_dataset.V7302, x = mtf_dataset.V7685)

    #Set x lable + xtick distances & lables
    ax.set(xlabel = col_names[1], xticks = ([1.4, 2.3, 3.15, 4, 4.8, 5.7, 6.6]), xticklabels = v7685_index)

    #Set hue question name and lables
    ax.legend(title = col_names[0],labels = v7302_index)

    plt.setp(ax.get_legend().get_texts(), fontsize='16')
    plt.setp(ax.get_legend().get_title(), fontsize='18')

    #plot histogram to subplot
    ax.plot()
    plt.show()

  #MTF heatmap
  def mtf_heatmap_full(mtf_col_names):

    plt.figure(figsize = (12, 9), dpi = 300)

    sns.heatmap(data = mtf_dataset.corr(), fmt = '.1f', linewidths = 1, cmap = 'Blues', annot = True,
xticklabels = mtf_col_names, yticklabels = mtf_col_names)


    plt.xticks(rotation=30)
    plt.yticks(rotation=30)


    plt.show()

  #-----------------------------------------------------------------------------------------------------------------#
  #Misc functions

  def yrbs_chi_squared_test(q80_index): # semi working produces output values
    #transform data for scipy.stats function
    yrx_data = pd.crosstab(yrbs_dataset['q25'], yrbs_dataset['q80']).T

    #scipy.stats - chi2 squared function
    c, p, dof, expected = sp.chi2_contingency(yrx_data)

    #Print results
    print('P-value: ', p)
    print('Test Statistic: ', c)
    print('Degrees of freedom: ', dof)
```

```python
def mtf_chi_squared_test(): # same as above

    #transform data for scipy.stats function
    mtf_data = pd.crosstab(mtf_dataset['V7302'],mtf_dataset['V7685']).T

    #scipy.stats - chi2 squared function
    c, p, dof, expected = sp.chi2_contingency(mtf_data)

    #Print results
    print('P-value: ', p)
    print('Test Statistic: ', c)
    print('Degrees of freedom: ', dof)



#-------------------------------------------------------------------------------------------------------------#
#menu functions

def print_menu(show_functions):
    print('-------------------------------------------------------------------------------')
    print('Welcome! \nThis program creates plots for parts of the YRBSS Youth risk survey \n& the
University of Michigan MTF survey')
    print('-------------------------------------------------------------------------------')

    print('\nEnter the number of plot to create:')
    print('#0: Back / Exit')
    for i in range(1, len(show_functions) + 1, 1):
        print(f'#{i}: {show_functions[i - 1]}')


#-------------------------------------------------------------------------------------------------------------#
#menu control logic

#main loop
while (True):

    #user input + variable type validation
    try:
        print_menu(show_functions)
        user_input = input('#: ')
        user_input = int(user_input)

        #If to  call menu option #0
        if (user_input == 0):
            break

        #elif to call menu option #1
```

```python
        elif (user_input == 1):
            print('Creating plot')
            yrbs_heatmap_full(yrbs_col_names)



        #elif to call menu option #2
        elif (user_input == 2):
            print('Creating plot')
            yrbs_stacked_histogram(yrbs_question_names, q80_index, q25_index)

        #elif to call menu option #3
        elif (user_input == 3):
            print('Creating plot')
            mtf_heatmap_full(mtf_col_names)

        #elif to call menu option #4
        elif (user_input == 4):
            print('Creating plot')
            mtf_stacked_histogram(mtf_question_names, v7685_index, v7302_index)

        #elif to call menu option #5
        elif (user_input == 5):
            print('Creating plot')
            yrbs_chi_squared_test(q80_index)

        #elif to call menu option #6
        elif (user_input == 6):
            print('Creating plot')
            mtf_chi_squared_test()

        #invalid input integer not in list
        else:
            print('\nThe entered value is invalid. Please enter the number of a plot:')

    except:
        print('\nThe entered value is invalid. Please enter the number of a plot:')

main()
```

Raw Data

To provide full transparency the datasets examined by this research can be found using the below link. **Please note** that while the YRBS [1] data as a work of the federal government is part of the public record, the MTF data [5] was evaluated with permission from the University of Michigan via a license from the National Addiction & HIV Data Archive Program and is not to be reproduced for further study without appropriate permissions.

https://github.com/Ds149/B104_Youth_Technology_Use-Behavioral_Health

Works Cited

[1] Centers for Disease Control and Prevention, "Youth Risk Behavior Survey Data," 2019.

[2] M. Underwood, N. Brener, J. Thornton, W. Harris, L. Bryan, S. Shanklin, N. Deputy, A. Roberts, B. Queen, D. Chyen, L. Whittle, C. Lim, Y. Yamakawa, M. Leon-Nguyen, G. Kilmer, J. Smith-Grant,, Z. Demissie, S. Everett Jones, H. Clayton and P. Dittus, "Overview and Methods for the Youth Risk Behavior," *US Department of Health and Human Services/Centers for Disease Control and Prevention - Morbidity and Mortality Weekly Report (MMWR) / Vol. 69 / No. 1,* 21 August 2020.

[3] U.S. Public Health Service, "Protecting Youth Mental Health, The U.S. Surgeon General's Advisory," 2021.

[4] M. Vuorre, A. Orben and A. Przybylski, "There Is No Evidence That Associations Between Adolescents' Digital Technology Engagement and Mental Health Problems Have Increased," *Clinical Psychological Science / Vol. 9(5) 823–835,* 2021.

[5] R. Miech, L. Johnston, J. Bachman, P. O'Malley, J. Schulenberg and M. Patrick, "Monitoring the Future: A Continuing Study of American Youth (8th- and 10th-Grade Surveys)," Inter-university Consortium for Political and Social Research [distributor], 2020-10-29, 2019.

[6] A. Lenhart and K. Owens, "Good Intentions, Bad Inventions: The Four Myths of Healthy Tech," Data & Society, 2020.

[7] P. Valkenburg and J. Peter, "The Differential Susceptibility to Media Effects Model," *Journal of Communication,* vol. 63, 2013.

[8] A. Lenhart and K. Owens, "The Unseen Teen: The Challenges of Building Healthy Tech for Young People," Data & Society, 2021.

[9] Pew Research Center, "Teens, Social Media & Technology May 2018," 2018.

Python Resources

https://matplotlib.org/stable/tutorials/index.html

https://www.w3schools.com/python/default.asp

https://pandas.pydata.org/docs/

https://numpy.org/doc/stable/

https://seaborn.pydata.org/

https://stackoverflow.com/