**Individual Project Proposals**

* Describe (a) the variables you plan to explore, and (b) the analysis you plan to perform
* Describe (a) the possible results your analysis may produce, and (b) how this could be relevant for the objectives of the course project
* Include your results if you have already completed an analysis

relationships with family, friends, coworkers, neighbours, and strangers -- for personal health and well-being. The purpose of our course project is to help raise awareness about and interest in this often underappreciated topic

With chatGPT’s help.

(And sorry I’m very sick during reading week… So I ask chatGPT to write my idea in detail and rephrase my words…)

For grouping:

(I don’t need to keep my research questions. I'm ok to do other topics. Please find me a group that wants to work well.. . plz plz plz plz plz…)

1:

**Research Question**

Does increased face-to-face interaction contribute to improved well-being among individuals?

**Variables**

**Independent Variable:**

**CONNECTION\_activities\_chat\_p3m**: In the PAST THREE MONTH, how often have you... - had an in-person, face-to-face conversation with someone?

This variable measures the frequency of in-person, face-to-face conversations individuals have had in the past three months. It is a categorical variable with responses such as "Never," "Rarely," "Sometimes," "Often," and "Very Often."

**Suggested Visualization**:

A histogram to display the frequency distribution of responses. This visualization would help identify the common levels of in-person interaction among participants, allowing us to see if most individuals have low, moderate, or high levels of face-to-face communication.

A bar plot would be useful for visualizing the distribution of responses. This visualization helps in understanding how often individuals engage in face-to-face interactions and allows for easy comparison of categories.

**Dependent variable:**

**WELLNESS\_phq\_score:** This score is derived from the Patient Health Questionnaire (PHQ-9) and measures levels of depressive symptoms. Lower scores indicate better mental health.

**Why choose this variable:**

The PHQ-9 (Patient Health Questionnaire) is a widely used tool for screening, diagnosing, and measuring the severity of depression. A lower PHQ-9 score indicates fewer depressive symptoms, which correlates with better mental health and overall wellness. This measure provides a quantitative assessment of an individual’s mental health status, allowing researchers to examine how face-to-face interactions may mitigate depressive symptoms.

**Suggested Visualization:** A box plot comparing PHQ scores across different categories of face-to-face interactions (e.g., low, medium, high). This visualization would help illustrate differences in well-being based on levels of in-person communication.

A box plot is an effective way to visualize the distribution of PHQ-9 scores across different categories of face-to-face interaction (e.g., low, medium, high).

**WELLNESS\_subjective\_happiness\_scale\_score:** This score is based on self-reported happiness, with higher scores indicating greater happiness.

**Why choose this variable:**

The subjective happiness scale measures an individual’s self-reported level of happiness and satisfaction with life. Higher scores indicate greater happiness, which is a key component of overall well-being. Happiness is linked to various positive health outcomes, including reduced stress, better immune function, and longer lifespan.

**Suggested Visualization:** A scatter plot with CONNECTION\_activities\_chat\_p3m on the x-axis and WELLNESS\_subjective\_happiness\_scale\_score on the y-axis, accompanied by a trend line. This visualization would help identify any correlation between face-to-face interaction and subjective happiness.

A scatter plot is ideal for showing the relationship between two continuous variables: face-to-face interaction frequency and happiness scores.

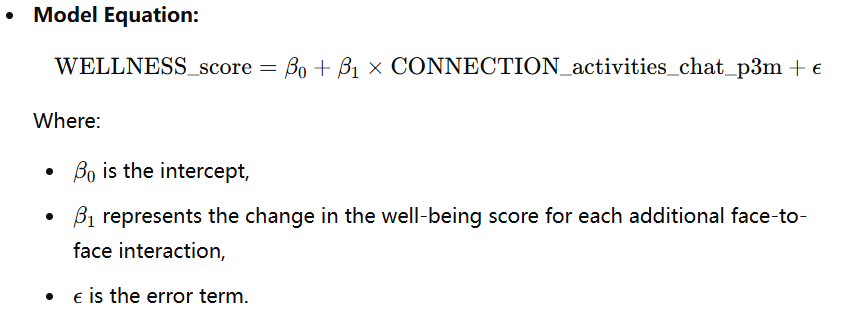
A scatter plot also allows for the addition of regression lines to quantify the relationship statistically. It provides a more nuanced understanding of how changes in one variable (face-to-face interactions) relate to changes in another (happiness).

**Controlled Variables:**Factors such as age, gender, socioeconomic status, and other social interactions should be controlled to isolate the effects of face-to-face communication on well-being.

**Analysis**

To analyze the relationship between face-to-face interaction and well-being, the following methods will be employed:

1. **Descriptive Statistics:**Begin by summarizing the data using mean, median, standard deviation, and frequency counts for both independent and dependent variables. This will help in understanding the general trends in the data.
2. **Data Visualization:**Use histograms and box plots to visualize the distribution of face-to-face interactions and well-being scores. This allows for a clear understanding of how these variables behave.
3. **Simple Linear Regression:**Fit a simple linear regression model with the frequency of face-to-face interactions as the independent variable and the well-being scores (either the PHQ score or the subjective happiness score) as the dependent variable. This model will help quantify the relationship between face-to-face interactions and well-being.



1. **Assumption Checks:**Check the assumptions of linear regression, including linearity (relationship between the variables should be linear), normality of residuals (the residuals should be normally distributed), and homoscedasticity (constant variance of residuals).
2. **Hypothesis Testing:**
   * **Null Hypothesis (H0​)**: There is no significant relationship between face-to-face interactions and well-being scores (β1​=0).
   * **Alternative Hypothesis (H1​)**: There is a significant positive relationship between face-to-face interactions and well-being scores (β1>0).
3. After fitting the regression model, examine the p-value associated with β1​.
   * **P-value Definition:** The p-value indicates the probability of observing the data (or something more extreme) assuming that the null hypothesis is true. A low p-value (typically less than 0.05) suggests that you can reject the null hypothesis, indicating a significant relationship between the variables.
4. **Confidence Intervals:**Construct confidence intervals for the slope coefficient β1\beta\_1β1​. This will provide a range of values within which we can be reasonably confident the true effect of face-to-face interactions on well-being lies.
   * **Interpretation:** If the confidence interval for β1\beta\_1β1​ does not include zero, this supports the conclusion that there is a significant relationship.

**Possible Results**

The analysis may reveal that individuals who engage in more face-to-face interactions report higher levels of well-being (lower PHQ scores and higher happiness scores). Such findings would highlight the significance of social connections for mental health, providing actionable insights for interventions aimed at improving social engagement.

**Hypothesis/results and discussion**

Increased frequency of face-to-face interactions positively correlates with improved well-being, as indicated by lower PHQ scores and higher subjective happiness scores.

If the results indicate a significant positive relationship, it would underscore the importance of maintaining social connections for enhancing individual well-being. This is particularly relevant in the context of our course project, as it raises awareness about the often underappreciated role of social interactions in promoting mental health.

2:

**Research Question**

How do preferences for spending time with different social groups (coworkers, family members, friends, and neighbors) relate to individual wellness?

**Variables**

**Independent Variables:**

**CONNECTION\_preference\_time\_coworkers\_classmates\_grouped**: How much time per week would you like to spend socializing with others from the following groups? - Coworkers or Classmates

Amount of time (in hours) individuals prefer to spend socializing with coworkers or classmates per week.

**CONNECTION\_preference\_time\_family\_grouped**: How much time per week would you like to spend socializing with others from the following groups? - Family Members

Amount of time (in hours) individuals prefer to spend socializing with family members per week.

**CONNECTION\_preference\_time\_friends\_grouped**: How much time per week would you like to spend socializing with others from the following groups? - Friends

Amount of time (in hours) individuals prefer to spend socializing with friends per week.

**CONNECTION\_preference\_time\_neighbours\_grouped**: How much time per week would you like to spend socializing with others from the following groups? - Neighbours

Amount of time (in hours) individuals prefer to spend socializing with neighbors per week.

**Suggested Visualization**:

**Bar Plots:**Create a bar plot for each of the four social groups (coworkers/classmates, family, friends, and neighbors) showing the average preferred time (in hours) individuals would like to spend socializing with each group. This will provide a clear comparison of how much time individuals prefer to spend with each group.

**Box Plots:**Use a box plot for each of the four social groups to display the distribution of preferred time spent. This will highlight the median, quartiles, and any potential outliers in the preferences for spending time with each group.

**Dependent Variables**: Wellness Scores

(may use the same scale as question 1, or other scales)

**Bar Plots:** Use bar plots to compare the average wellness scores (PHQ and subjective happiness) across different social group preferences. This visualization will help highlight any trends or differences in wellness associated with the time spent with each social group.

**Box Plots:** Create box plots for each social group's preferred time to visualize the distribution of wellness scores. This will aid in identifying outliers and the central tendency of wellness scores within each group.

**Analysis**

To explore the relationship between preferences for spending time with different social groups and individual wellness, the following methods will be applied:

1. **Descriptive Statistics:**Summarize the data for each independent variable (time spent with coworkers/classmates, family, friends, and neighbors) and the dependent wellness variables (PHQ score and subjective happiness score). Calculate means, medians, standard deviations, and frequency distributions.
2. **Data Visualization:**
   * **Bar Plots:** Create bar plots for each social group (coworkers/classmates, family, friends, neighbors) to show the average preferred time (in hours) individuals would like to spend socializing. This provides a visual comparison of preferences across groups.
   * **Box Plots:** Use box plots to display the distribution of preferred time for each social group, highlighting the median, quartiles, and any potential outliers. This visualization will help identify variations in preferences.
3. **Correlation Analysis:**Perform correlation analysis (e.g., Pearson or Spearman correlation) to assess the relationship between the amount of time individuals prefer to spend with each social group and their wellness scores. This will help determine whether there is a positive or negative association between social preferences and well-being.
4. **Hypothesis Testing:**For each social group, we will formulate null and alternative hypotheses to test if preferences for spending time with that group are significantly related to wellness scores.
   * **Null Hypothesis (H0​)**: There is no significant relationship between the preferred time spent with [social group] and wellness scores (no correlation).
   * **Alternative Hypothesis (H1​)**: There is a significant relationship between the preferred time spent with [social group] and wellness scores (there is a correlation).
5. After calculating the correlation coefficient, we will determine the p-value associated with this correlation.
   * **P-value Definition:** The p-value indicates the probability of observing the data (or something more extreme) assuming that the null hypothesis is true. A low p-value (typically less than 0.05) suggests rejecting the null hypothesis, indicating a significant relationship between the time spent with that social group and individual wellness.
6. **Confidence Intervals:**Construct confidence intervals for the correlation coefficients obtained from the correlation analysis. This will provide a range of plausible values for the correlation, allowing for a better understanding of the strength and direction of the relationship.
   * **Interpretation:** If the confidence interval for the correlation coefficient does not include zero, it indicates a significant relationship between the preferred time spent with that social group and wellness.

**Possible Results**

The analysis may reveal that preferences for spending time with certain social groups are significantly correlated with wellness scores. For instance, if individuals who prefer to spend more time with family or friends report higher happiness scores, this highlights the importance of social connections for well-being.

Such findings can provide insights into how spending time with different social groups contributes to individual health and well-being, informing future interventions or public health campaigns aimed at fostering social interactions for better mental health.

**Hypothesis/results and discussion**

The analysis aims to determine whether preferences for spending time with various social groups significantly relate to individual wellness. If the hypothesis tests yield significant results, we may find that individuals who prefer to spend more time with specific groups, such as family or friends, report better wellness outcomes (lower PHQ scores and higher happiness scores).

* **Expected Findings:**
  + It is anticipated that spending more time with family and friends may be associated with better well-being compared to time spent with coworkers or neighbors, reflecting the importance of social support in mental health.
  + These findings will underscore the relevance of social interactions for enhancing wellness, informing personal health strategies and community initiatives aimed at fostering stronger social connections.

3:

**Research Question**

How does physical contact, such as hugging and kissing, influence individual wellness?

**Variables**

**Independent Variables:**

**CONNECTION\_activities\_hug\_p3m**: In the PAST THREE MONTH, how often have you... - hugged someone?

Frequency of hugging someone in the past three months (measured on a Likert scale, e.g., never, rarely, sometimes, often, very often).

**Visualization Suggestion:** A bar plot showing the distribution of responses for hugging frequency. This visualization will help identify how often individuals engage in hugging, providing a clear picture of the data distribution.

**CONNECTION\_activities\_kissed\_p3m**: In the PAST THREE MONTH, how often have you... - kissed someone?

Frequency of kissing someone in the past three months (also measured on a Likert scale).

**Visualization Suggestion:** Similar to hugging, a bar plot for the frequency of kissing will visualize the distribution of responses. This will allow for easy comparison between hugging and kissing frequencies among respondents.

**Dependent Variables:**

**WELLNESS\_phq\_score:** A score derived from the Patient Health Questionnaire assessing mental health and well-being.

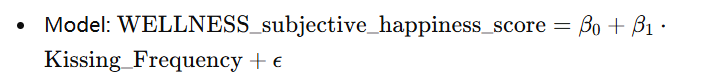
**Visualization Suggestion:** A box plot representing the distribution of PHQ scores based on different frequency categories of hugging and kissing. This will highlight any differences in wellness scores across various levels of physical contact.

**WELLNESS\_subjective\_happiness\_scale\_score:** A score reflecting individuals’ self-reported happiness.

**Visualization Suggestion:** Similar to the PHQ score, a box plot for happiness scores based on hugging and kissing frequencies will help illustrate the relationship between these physical contacts and subjective happiness levels.

**Analysis**

To investigate the influence of physical contact (hugging and kissing) on individual wellness, the following methods will be implemented:

1. **Descriptive Statistics:**Calculate descriptive statistics for the independent variables (frequency of hugging and kissing) and the dependent wellness variables (PHQ score and subjective happiness score). This includes means, medians, standard deviations, and frequency distributions to provide an overview of the data.
2. **Data Visualization:**
   * **Bar Plots:** Create bar plots for both hugging and kissing frequencies to show the distribution of responses. This visualization will help identify how often individuals engage in these physical contacts and provide a clear comparison between the two.
   * **Box Plots:** Use box plots to display the distribution of PHQ scores and subjective happiness scores based on different frequency categories of hugging and kissing. These visualizations will highlight any potential differences in wellness scores across various levels of physical contact.
3. **Correlation Analysis:**Assess the correlation between the frequency of hugging and kissing and the wellness scores using Pearson or Spearman correlation coefficients, as appropriate. This will indicate whether there is a relationship between the physical contact frequencies and individual wellness.
4. **Hypothesis Testing:**For both hugging and kissing, we will formulate null and alternative hypotheses to test the relationship between each type of physical contact and the wellness scores.
   * **Null Hypothesis (H0)**: There is no significant relationship between the frequency of hugging and wellness scores (PHQ score and happiness score) (no correlation).
   * **Alternative Hypothesis (H1​)**: There is a significant relationship between the frequency of hugging and wellness scores (PHQ score and happiness score) (there is a correlation).
   * **Null Hypothesis for Kissing:** Similar formulation as for hugging.
5. After calculating the correlation coefficients, we will determine the p-values associated with these correlations.
   * **P-value Definition:** The p-value indicates the probability of observing the data (or something more extreme) under the null hypothesis. A low p-value (typically < 0.05) suggests that we can reject the null hypothesis, indicating a significant relationship between physical contact and wellness scores.
6. **Simple Linear Regression Analysis:**Conduct two separate simple linear regression analyses:
   * **First Regression Analysis:** Assess the relationship between the frequency of hugging (independent variable) and wellness scores (dependent variable: PHQ score).
   * **Second Regression Analysis:** Assess the relationship between the frequency of kissing (independent variable) and wellness scores (dependent variable: happiness score).
7. In each analysis, interpret the coefficients (β1) to determine how changes in hugging or kissing frequency are associated with changes in wellness scores. Also, check R-squared values to assess how well the models explain the variability in the wellness scores.
8. **Confidence Intervals:**Construct confidence intervals for the regression coefficients obtained in each simple linear regression analysis. This will provide a range of plausible values for the relationship, allowing a clearer understanding of the influence of physical contact on wellness.
   * **Interpretation:** If the confidence interval for the regression coefficient does not include zero, it indicates a significant relationship between physical contact and wellness.

**Possible Results**

The analysis may reveal that higher frequencies of hugging or kissing are associated with better wellness scores (lower PHQ scores and higher happiness scores). For example, if individuals who hug more frequently report higher happiness scores, this finding underscores the importance of physical contact for mental health and well-being.

Such insights can enhance awareness of the positive effects of physical affection and inform public health initiatives aimed at promoting healthy interpersonal relationships for improved mental well-being.

**Hypothesis/results and discussion**

The analysis aims to determine whether physical contact, such as hugging and kissing, significantly influences individual wellness. If the hypothesis tests yield significant results, we may find that higher frequencies of hugging and kissing are associated with better wellness outcomes (lower PHQ scores and higher happiness scores).

* **Expected Findings:**
  + It is anticipated that individuals who hug and kiss more frequently report better wellness outcomes, suggesting that physical contact has a positive impact on mental health and happiness.
  + These findings will highlight the importance of physical touch in fostering emotional well-being, which can be used to encourage greater physical interactions in personal relationships and community initiatives focused on mental health.