1. Factors that influence level of loneliness and personal health

Reserch question: What influence the level of loneliness and personal mental health(e.g.age, region, gender) which contributes to a higher rates of morbidity and how far do people feel lonely as measured by the data collected through questionnaire

Variables:

Dependent Variables:

·LONELY_ucla_loneliness_scale_score

Description: a numeric score representing to what certain extent that people feel lonely out of a total score of 10 based on their responses to a specific questionnaire

Visualization: Histogram

Reasoning: A histogram can showcased the general distribution of people with different level of loneliness. Additionally, it's clear to see where the central trends locates. Statistics can also indicated in a clear way through Histogram such as median and mean. Some special individuals can also be included in Histogram through outliers and skewness of Histogram.

·LONELY_dejong_emotional_loneliness_sub_scale_score

Description: a numeric score up to 3.0 representing to what extent people feel emotional lonely based on their responses to a specific questionnaire

Visualization: Box Plot

Reasoning: A box plot can perfectly showcased the summary of the distribution of the data through the median, quartiles and further allowing for comparisons between different demographic groups. It also highlights the skewness of the data and where the outliers are, which is very helpful for visualizing variability and central tendency.

Independent variables:

·DEMO_age

Description: A numeric variable showcasing the age of each individual in years

Visualization: Scatter Plot
Reasoning: Scatter plots allow us to find relationship between two numeric variables and we can

see the general trends and correlations between age and level of loneliness very clearly, demonstrating how far does age affect people's level of loneliness.

·DEMO_gender

Description: A categorical variable indicating the gender of the individual such as male, female or non-binary

Visualization: Bar Chart

Reasoning: A bar chart is pretty straightforwaard for indicating the differences across genders of their mean level of loneliness, indicating whether gender experience a different level of loneliness

·DEMO_identity_indigenous, DEMO_identity_disability, DEMO_identity_homeless

Description: Variables of special individual groups indicating people with special conditions like being homeless or disabilities

Visualization: grouped bar chart

Reasoning: grouped bar charts can easily provide a clear comparison of Indigenous people, disabilities and homeless people, indicating the difference between the level of loneliness according to their backgrounds and conditions

Analysis:

-Methodology:
a) Hypothesis Testing:
Set up assumptions(H_0 and H_1), then use t-test to calculate the statistics in order to further calculate the p-value, then choose a significant level, which is normally 0.05, compare the p-value with the significant level and eventually either reject H_0 or fail to reject H_0
b) Simple Linear Regression:
Use the collected data in the questionnaire and fit the regression model. Then interpret the coefficient such as the slope, which indicates the change in Y for a one-unit increase in X. A positive slope means Y increases as X increases, while a negative slope means Y decreases. And the intercept, which represents the predicted value of Y when X is 0. Also check the p-value for the slope, if the p-value is less than your significance level (e.g., 0.05), it suggests that the relationship between X and Y is statistically significant.

Hypothesis:
Based on the data showcased by the questionnaire, we can predict that nowadays the society do have the issue of experiencing loneliness and huge amounts of people are currently struggling with loneliness. Also, people's sense of lonely is pretty strong, as most of the people have high level of loneliness(around 6 to 9 when out of 10) while some others spread at a score of 4. Seldom people

have very low level of loneliness. Furthermore, we anticipate that due to social issue like aging of population and social stress, it will have a positively correlation between age and level of loneliness. However, gender is not a big factor for affecting level of loneliness since all genders tend to suffer from loneliness and loneliness is not relevant to genetic issues. Nevertheless, we still cannot rule out the possibility of the difference between level of sensitivity due to genders. Moreover, regions will affect the level of loneliness to some extent since the population density can directly influence the happiness of the local people.

Relevant for the objectives of the course project:
Through investigation of factors that affect people's loneliness, we can seriously realize one of the biggest issue in society, which is loneliness. Also, we can find out some ways to heal the mental scar brought by loneliness in society, which further increase the level of happiness. Moreover, we can take a deeper insight into how social issue can influence each individual's mental health and self-awareness, which help promote mental health advantages and improve personal satisfaction for different groups with different ages and regions.

2. The health degree of people in society from both physical and mental perspective

Reserch question: Whether or not today's people are healthy both physically and mentally, if not, what caused their physical or mental depression

Variables:
Dependent Variables:

·WELLNESS_self_rated_physical_health

Description: A numeric score indicating how far do people think they are physically healthy, the scores are divided into several parts such as good, fair, very good and so forth.

Visualization: Histogram

Reasoning: Histogram can clearly showcase the general distribution of people's extent of physical health, indicating the central trend with special individuals as outliers. Through Histogram, we can also view the mean and median, which help us get a better realization of society's overall physical health level.

·WELLNESS_self_rated_mental_health

Description: A numeric score indicating how far do people think they are mentally healthy, the scores are also divided into several parts showcased through numbers(e.g.5,6,7,3)

Visualization: Histogram also/ box plot

Reasoning: Same as physical ones, histogram can show the central trend of distribution in a clear way with outliers being indicated. Also, box plot is pretty helpful as well since it can highlight the skewness and the outliers while showcase the quartiles and median of the distribution, we can make a comparison between mental health distribution and physical ones through these methods.

Independent Variables:

・*DEMO_age*

Description: A numeric variable showcasing the age of each individual in years

Visualization: Scatter Plot

Reasoning: Scatter plots allow us to find relationship between two numeric variables and we can see the general trends and correlations between age and level of physical health and mental health very clearly, demonstrating how far does age affect people's health degree.

・DEMO_gender:

Description: A categorical variable indicating the gender of the individual such as male, female or non-binary

Visualization: Bar Chart

Reasoning: A bar chart is pretty straightforwaard for indicating the differences across genders of their mean health degrees, indicating whether gender experience a different health degree

・DEMO_identity_disability

Description: A text description of people's body condition

Visualization: (grouped) bar plots

Reasoning: we can use grouped bar plots to analyse the how disability type, severity, age, income, and other factors predict physical and mental health outcomes, based on the distribution of the factors and health degrees, we are able to focus more on health degrees with disabilities.

Analysis:

Methodology:
Multiple Linear regression:we can use multiple linear regression to provide a comparison of the degree of health with several factors from both mental and physical perspective. Through the correlation of each factor, we can define the severity of each factor, whether the factor has a positive relation with health, which is very helpful for analyzing several factor at the same time.

Hypothesis:
We infer that nowadays, people's condition of physical health is better than their mental health based on the data in the questionnaire. Most of the people are physically good as they said. However, their mental health degree is as high as their physical ones. Additionally, we speculate that factors like gender and age do not matter too much as the degree of health is pretty random for different ages and genders.

Relevant for the objectives of the course project:
Through this investigation of people's self-awareness to their own health degrees, we can have a deeper understanding of people's physical and mental condition. As people now suffering from social stress, peer pressure and expectations from family, the depression brought by work and school is harsh. After realizing this, we can do something to help provide a healthier way of competition in society, or offer some mental institution to help people solve their mental problems, which can further promote the creativity and societal efficiency of production due to increase of people's satisfaction of the society and health degrees.

3. Factors that affect the possibility of getting COVID-19

Research question: what are the factors that affect the possibility of getting COVID-19 and to what extent does each factor influence the possibility

Variables:
dependent variables:

·COVID_prevention_distancing, COVID_prevention_masks，COVID_prevention_avoid_trips, COVID_prevention_reduce_people

Description: A scale of different degree of relations with distance between people, wearing masks, avoiding trips and reduce population density and the possibility of getting COVID-19(very closely, somewhat closely and so on)

Visualization: Histogram

Reasoning: Histogram can showcase the overall distribution of different extent of how distance and wearing masks relates to the possibility of getting COVID-19 in people's opinion. Furthermore, we can easily see the central trend of people's thoughts with mean and median. The outliers can be indicated as well, which is pretty useful.

Independent Variables:

· *DEMO_age*

Description: A numeric variable showcasing the age of each individual in years

Visualization: Scatter Plot

Reasoning: Scatter plots allow us to find relationship between two numeric variables and we can see the general trends and correlations between age and level of physical health and mental health very clearly, demonstrating how far does age affect people's health degree.

・DEMO_gender:

Description: A categorical variable indicating the gender of the individual such as male, female or non-binary

Visualization: Bar Chart

Reasoning: A bar chart is pretty straightforwaard for indicating the differences across genders of their mean health degrees, indicating whether gender experience a different health degree

Analysis:

Methodology:
Confidence Interval:for each extent of how each factor affect the possibility of getting COVID-19, we create a confidence interval, after collecting the data, we calculate the sample proportion of each factor. Then, we set the confidence level and calculate the confidence interval and interpret the intervals. Finally, we compare each interval and see which is most widely followed. We can also consider the width of the interval, if the interval is narrow, the sample proportion is stable, if it's wide, there's more variability.

Hypothesis:
Through this investigation, we can infer that each of the factors included in this investigation has strong relationship with the possibility of getting the COVID-19 since the disease is highly spreadable and can spread in a very rapid speed. So we anticipate that every single factor has a positive relation with the possibility but with different extent.

Relevant for the objectives of the course project:
As we investigate on the factors that affect the possibilities of getting COVID-19, we can speculate that COVID-19 is highly spreadable as every single movement and policy is strongly related to the final possibility. This investigation help us have a better awareness of the importance of maintaining a good personal hygiene in our daily life and give an alert to the society that we should pay more attention to COVID-19 and make some contributions to the society thorough taking daily precautions and have the evil thought of bringing other people troubles.

Last but not least, for group project, I want to corporate with Jade Wang and Harry Meng. Because firstly we're at the same tutorial class and we often discuss our own thoughts of different kinds of statistics issues during tutorial classes. Secondly, we can distribute the work very clear

and appropriate when processing group works, every one of us would contribute equally to the project. Additionally, we're all willing to help each other when facing problems and we always can find a solution.