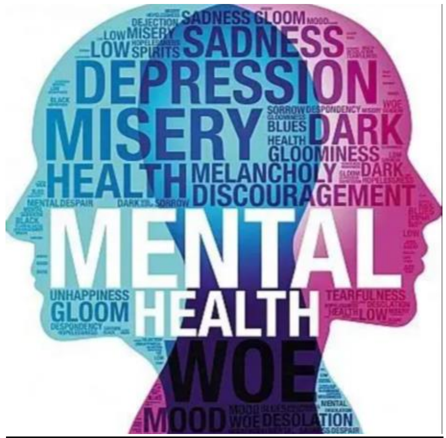
DSA4262 Individual Assignment 1
(Tan Teck Hwe Damaen, A0259079M)



Brief Overview Of The Project And Analysis:

In recent years, depression has emerged as a prominent public health concern in Singapore, receiving increased attention from policymakers, healthcare institutions and the media. Rising discussions around youth mental health, academic pressure and the pervasive use of social media have highlighted depression as a growing and complex issue, particularly among younger populations. This heightened focus underscores the need for evidence-based analysis to better understand who is most affected and where interventions are the most effective.

Hence, for individual assignment 1 on data visualization, I have focused on the area of mental health, specifically depression. The study aims to identify populations at the highest risk of depression, examine the most significant contributing factors and inform the development of targeted policies to help reduce depression rates in Singapore.

Datasets Used And Why They Are Used:

For the purpose of this assignment, I have made use of **3 datasets** from various websites such as Kaggle, Institute For Health Metrics And Evaluation (IHME) and data.gov.sg.

Note: There are **2 additional datasets** from Our World In Data and Kaggle used for the full project, but these datasets will not be listed in this assignment due to the restrictions of 3 distinct visualizations. Kindly refer to the full analysis of the project in the GitHub repository, where I have published the Jupyter Notebook containing the full project details. The full project includes data cleaning and data visualization. Code for data cleaning and data visualization is performed in Python. (Link To Full Project: GitHub Repository Individual Assignment 1 Jupyter Notebook)

Introducing The 3 Datasets:

Dataset 1: Depression DALY Contribution By Age Group And Gender [Link To Dataset]

Why Is This Dataset Used?

This dataset provides a comprehensive measure of the burden of depression across different demographic groups including age and gender. By quantifying the Disability-Adjusted Life Years (DALYs) attributable to depression, the dataset allows for a nuanced understanding of which age groups and genders are most affected. This information is critical for identifying high-risk populations, assessing the relative impact of depression across demographics and informing evidence-based policy interventions in Singapore subsequently. Using this dataset ensures that the analysis is grounded in reliable, population-level health metrics, enhancing accuracy and relevance of the study's findings and recommendations.

Dataset 2: Anxiety And Depression Mental Health Factors [Link To Dataset]

Why Is This Dataset Used?

This dataset is selected because it provides comprehensive information on multiple factors associated with anxiety and depression across different demographic groups. Its detailed coverage of potential causes such as sleep hours, physical activity and family history enables a thorough analysis of which variables contribute most significantly to depression. By linking these factors to demographic characteristics, the dataset allows us to identify the root causes of depression for the most vulnerable demographics. Additionally, through this dataset, we can conduct correlation analysis between depression levels and the contributing factors, providing actionable insights that can inform the development of targeted mental health policies and interventions in Singapore.

Dataset 3: Singapore Resident Population by Planning Area/Subzone, Age Group and Sex (General Household Survey 2015) [Link To Dataset]

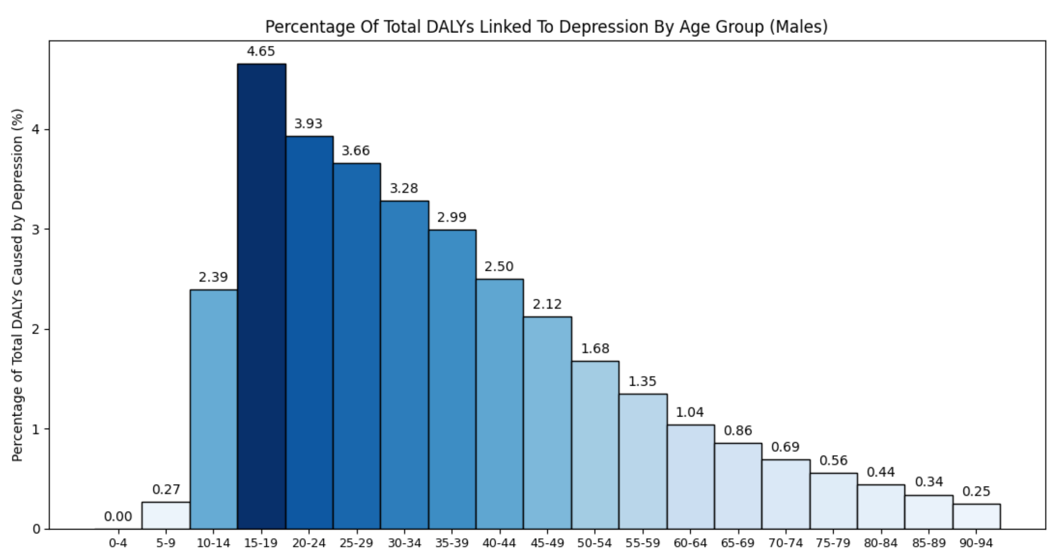
Why Is This Dataset Used?

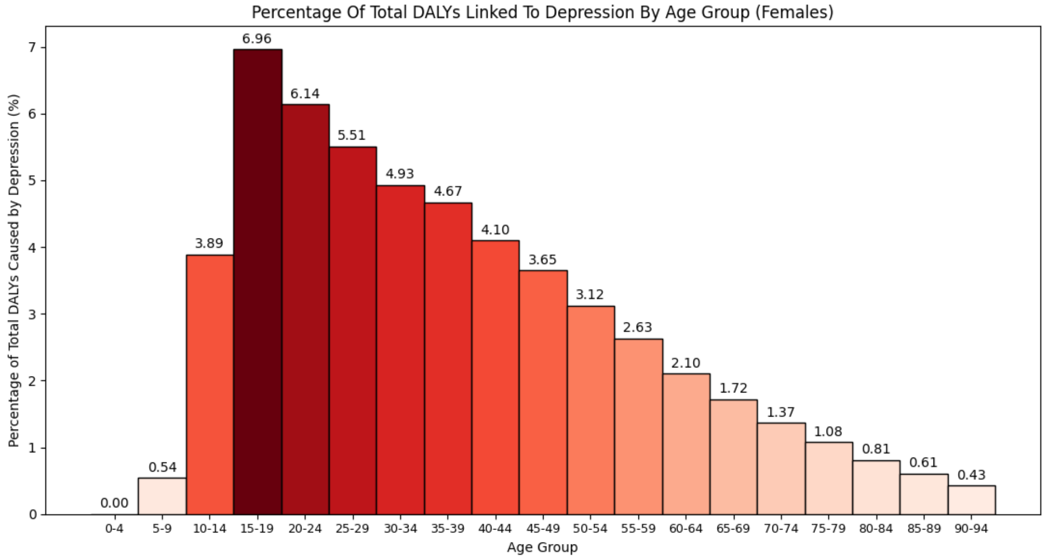
The dataset provides detailed demographic information on Singapore's resident population, segmented by planning area/subzone, age group and gender. It allows for the identification of specific estates where vulnerable populations based on age and gender are concentrated. By analyzing the distribution of these high-risk groups, policymakers and public health authorities can implement more targeted interventions, such as community mental health programs in areas with the highest prevalence of at-risk populations. Using this dataset ensures that the analysis is accurate in the context of Singapore, enabling data-driven decision making for depression prevention and mitigation efforts in Singapore.

Plot 1: The Macro Plot

For this visualization, I used **Dataset 1: Depression DALY Contribution By Age Group And Gender** to explore depression rates across different demographic groups. The goal is to examine how depression levels vary by age group and gender, providing a broad perspective on the populations most affected. This plot is considered a macro plot because it captures a high-level, population-wide view rather than focusing on individual cases or localized subsets. By analyzing the graph, we aim to identify which age groups and gender cohorts experience the highest rates of depression, thereby highlighting where public health interventions may be most needed.

The metric used, Disability-Adjusted Life Years (DALY), measures the overall burden of a disease by combining years of life lost due to premature mortality with years lived with disability. Specifically, the DALY contribution for depression reflects the impact of depressive disorders on population health, providing a quantitative way to gauge the prevalence and severity of depression. This metric is particularly useful because it allows us to compare depression levels across age groups and gender in a standardized manner, offering insights that can guide targeted mental health strategies subsequently.





Plot 1: Percentage Of Total DALYs Linked To Depression By Age Group

Plot 1 consists of two subplots – the top subplot represents males, and the bottom subplot represents females. Both plots show the percentage of total Disability-Adjusted Life Years (DALYs) linked to depression across different age groups. The plots use color-blind friendly gradients (shades of blue for males and shades of red for females). In each subplot, darker shades indicate higher percentage of DALYs linked to depression, providing an intuitive visual cue for identifying the age groups most affected by depression. This subplot format allows for a clear gender comparison while maintaining consistency in age group representation, making it easier to identify trends across the lifespan.

Analysis And Observations:

From the visualization:

1. Peak depression burden in adolescence and young adulthood
   - For **males**, the highest DALY depression percentage occurs in the **15-19** age group (4.65%), followed by 20-24 (3.93%) and 25-29 (3.66%)
   - For **females**, the peak is similarly observed in **15-19** (6.96%), followed by 20-24 (6.14%) and 25-29 (5.51%).
   - This indicates that **depression exerts the greatest health burden during the teenage and early adult years** for both genders

2. Lower DALY depression contribution in early childhood and older age
   - Children under 10 years show negligible depression-related DALYs
   - Adults over 50 years display a gradual decline in DALYs

3. Gender differences
    - Across all age groups, **females consistently show higher percentages of DALYs linked to depression** compared to males
    - This aligns with the global trends indicating that females are more likely to experience depression potentially due to biological, hormonal and psychosocial factors

Several factors may explain the observed trends. Firstly, adolescents involve rapid hormonal changes, which may increase vulnerability to mood disorders in both males and females. Moreover, teenagers and young adults face academic pressure, social identity development and career uncertainty, contributing to higher depression burden. Individuals in their late teens to early 20’s experience major life transactions such as tertiary education, entering the workforce and forming independent social networks, which may increase stress and susceptibility to depression. Lastly, females may experience higher rates of depression compared to males due to factors such as hormonal fluctuations, societal expectations and higher likelihood of reporting symptoms.

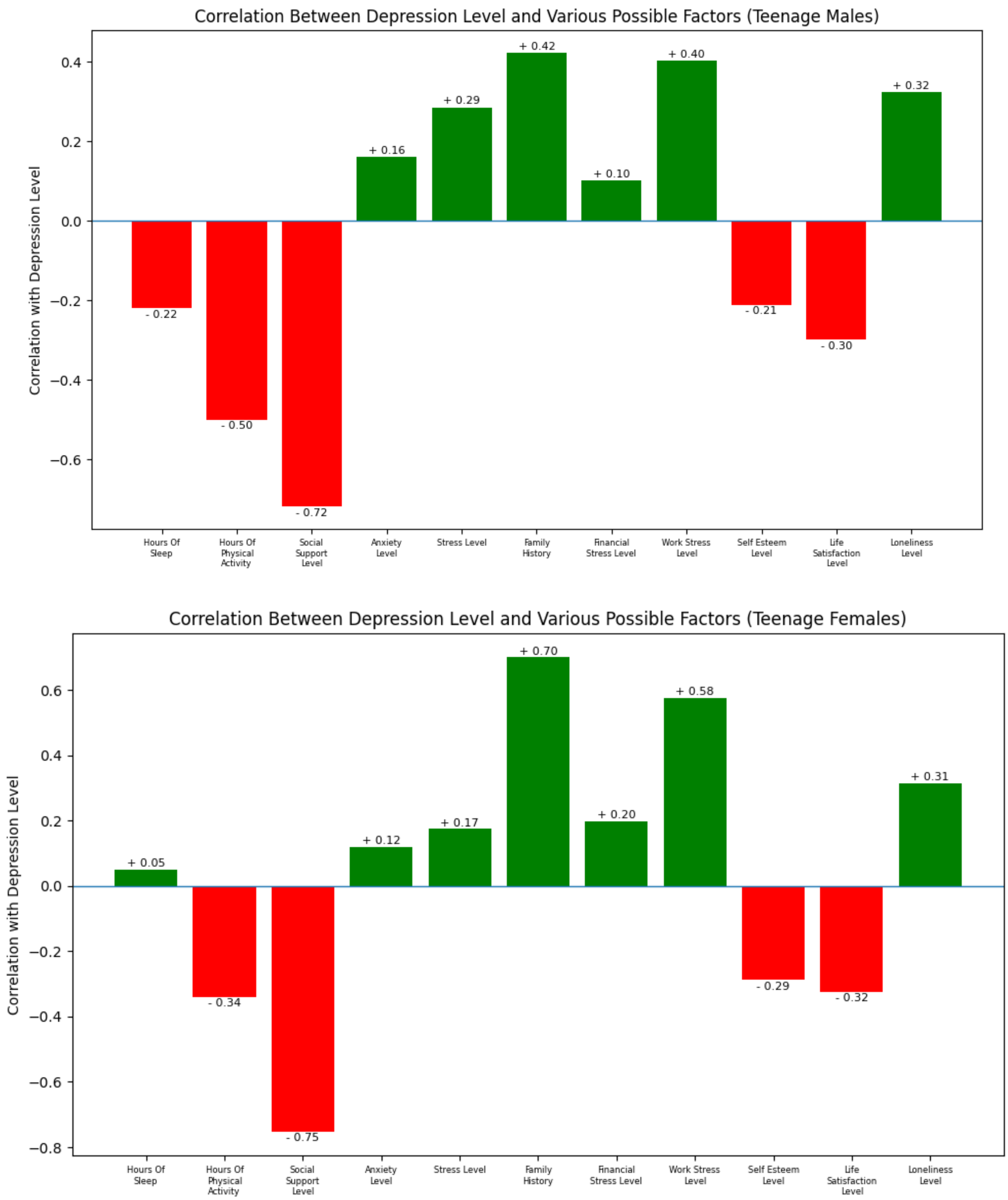
However, we should not immediately jump to conclusions. We need to dive deeper into the possible factors of depression, and which factors are most involved in causing depression in teenage females and males. This is when Plot 2 comes into action.

Plot 2: The Micro Deep Dive Plot

Plot 2 serves as a micro deep-dive plot as it moves beyond population level patterns to examine the underlying factors contributing to depression within a specific high-risk group – teenagers aged 15 to 19. Unlike a macro plot that establishes broad tasks across age and gender, this plot focuses on why depression is more prevalent in this demographic by identifying potential root causes.

Using **Dataset 2: Anxiety And Depression Mental Health Factors**, the analysis examines the relationships between depression levels and various contributing factors through correlation analysis. The dataset includes multiple variables associated with mental health such as anxiety, sleep patterns, physical activity and other psychological factors, allowing these to be treated as exploratory variables. By visualizing the correlations between these factors and depression levels, the plot highlights which factors are most strongly associated with depression among teenagers, offering insights into the mechanisms driving the elevated depression burden observed in this age group. This micro-level analysis provides actionable evidence to inform targeted interventions aimed at addressing the most influential drivers of teenage depression.

These correlations are exploratory and intended to generate hypotheses rather than establish causal relationships.



Plot 2: Correlation Between Depression Level And Factors Of Depression

Plot 2 is presented with 2 subplots – with the top subplot representing teenage males aged 15-19 and the bottom subplot representing teenage females aged 15-19. This layout enables direct gender-based comparisons while maintaining consistency across factors. Each subplot displays a bar chart of correlation coefficients between depression levels and various contributing factors. A color-blind friendly converging color scheme is used to clearly distinguish correlation direction.

- **Green bars** indicate **positive** correlation, where higher levels of a factor are associated with higher depression levels.
- **Red bars** indicate **negative** correlations, where higher levels of a factor are associated with lower depression levels.
- The horizontal reference at correlation = 0 helps visually separate positive and negative relationships.
- Numerical correlation values are annotated directly on each bar, improving interpretability and allowing for a quick comparison of correlation magnitudes, which is the primary focus of the analysis.

From the plot obtained, almost all of the directions of correlations aligns with theoretical expectations, reinforcing the validity of the dataset (for the full explanation and the anomaly observed, please refer to the Jupyter notebook uploaded on GitHub [GitHub Repository Individual Assignment 1 Jupyter Notebook]). Protective factors such as social support, physical activity and life satisfaction exhibit negative correlations with depression while stress-related and vulnerability factors show positive correlations.

However, rather than just focusing on the direction of correlation, this analysis emphasizes the magnitude of correlations, as stronger magnitudes indicate factors more closely associated with changes in depression levels.

Analysis And Observations:

From the visualization:

1. For **Teenage Males**, the four factors with the highest depression correlation magnitude are
    - **Social Support Level** (|r| = 0.72)
    - **Hours Of Physical Activity** (|r| = 0.50)
    - **Family History** (|r| = 0.42)
    - **Work Stress Level** (|r| = 0.40)

   These results suggest that both social-environmental factors (social support, family history) and lifestyle or stress-related factors (physical activity) play a significant role in influencing depression levels among teenage males.

2. For **Teenage Females**, the four factors with the highest depression correlation magnitude are
    - **Social Support Level** (|r| = 0.75)
    - **Family History** (|r| = 0.70)
    - **Work Stress Level** (|r| = 0.58)
    - **Hours Of Physical Activity** (|r| = 0.34)

   Compared to males, teenage females exhibit stronger correlations for social support and family history, suggesting heightened sensitivity to interpersonal and familial contexts.

3. A key insight from the micro-level analysis is that the **same four factors (Social Support Level, Family History, Work Stress Level and Hours Of Physical Activity) emerge as the most significant contributors to depression for both teenage males and females**. The primary difference lies in the relative strength of these associations, with females generally showing stronger correlations, particularly for social and familial factors. This reinforces findings from psychological literature that females may be more vulnerable to interpersonal stressors and relational dynamics, while males exhibit significantly balanced influences across social, lifestyle and stress-related domains.

The strength of the correlations can be explained by the developmental and psychosocial characteristics of adolescence.

- **Social Support Level**: Adolescence is a critical stage for emotional development, peer validation and identity formation. Limited social support can directly intensify feelings of isolation, rejection and low self-worth – core components of depressive symptoms. This explains why social support exhibits the strongest correlation for both genders.

- **Family History**: Family history reflects both genetic predisposition and shared environmental influences, such as parenting styles, emotional availability and household stress. Teenagers, who have limited independence, are particularly affected by familial environments, leading to stronger associations with depression.

- **Hours Of Physical Activity:** Physical activity is closely linked to mood regulation, stress level and sleep activity. Modern academic pressures and increased screen time reduce physical activity among teenagers, making inactivity a significant contributor to depressive symptoms.

- **Work Stress Levels:** For teenagers, work stress often represents academic workload, examination pressure, part-time employment and future uncertainty. These stressors are chronic and high stakes, while teenagers typically lack mature coping mechanisms, resulting in a strong association with depression levels.

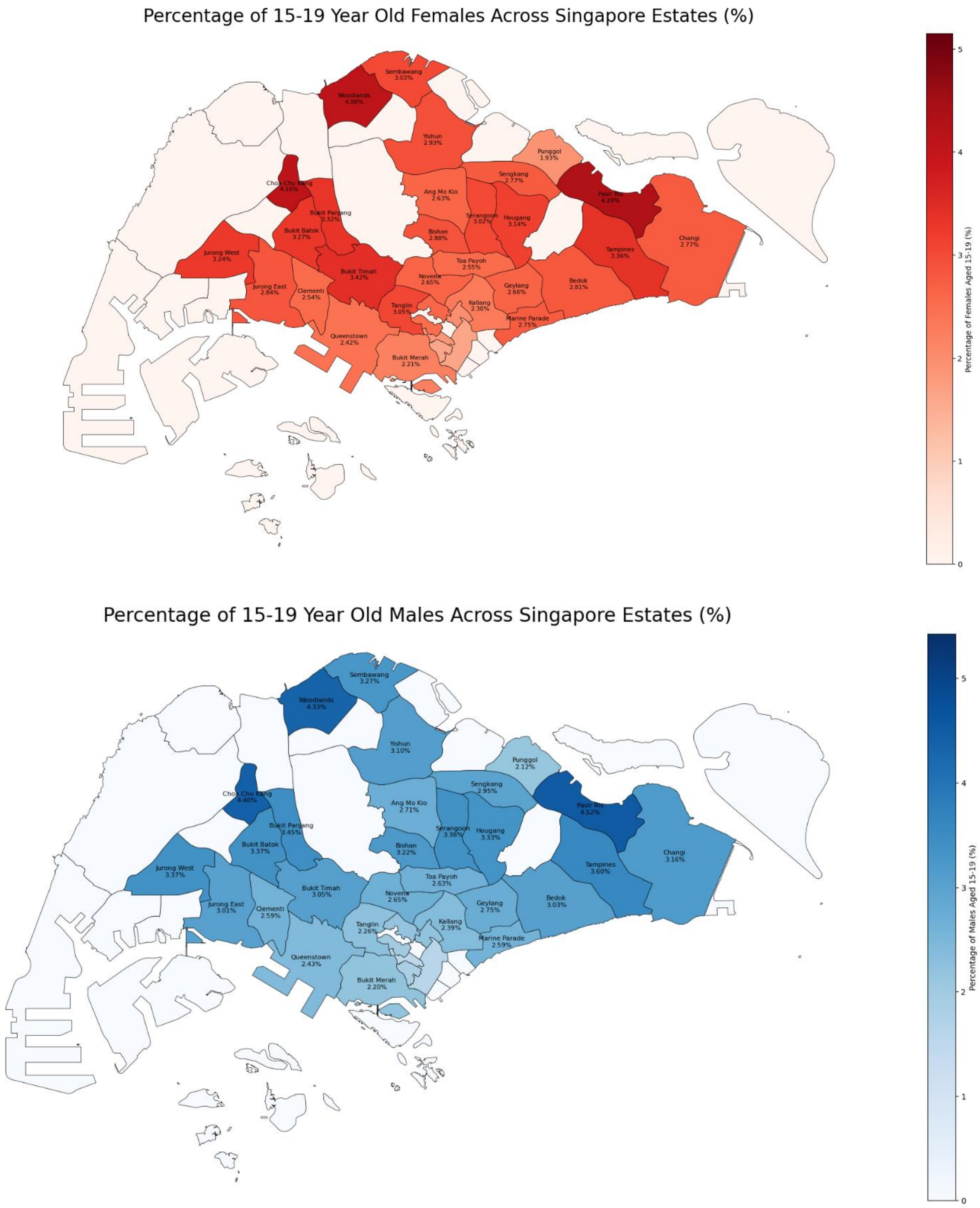
With the key root causes of depression among teenage males and females aged 15 to 19 identified through Plot 2, the analysis can now progress toward the formation of targeted strategies aimed at reducing depression levels in Singapore. In particular, the findings highlight that Social Support Level, Family History, Hours Of Physical Activity and Work Stress Levels are the factors most strongly correlated with depression during adolescence.

While these insights provide a clear direction for policy design, effective implementation requires an additional layer of information. To ensure that interventions are both targeted and impactful, it is essential to first identify where teenagers aged 15 to 19 are geographically concentrated across Singapore. Understanding the spatial distribution of this high-risk demographic will enable policymakers to deploy resources, programs and support services in areas where they are most needed, thereby increasing the effectiveness of proposed mental health interventions. This is when Plot 3 comes into action.

Plot 3: The Actionable Plot

This visualization serves as an actionable plot because it translates analytical insights into clear geographic targets for policy intervention. Using **Dataset 3: Singapore Resident Population By Planning Area/Subzone, Age Group And Sex (General Household Survey 2015)**, the plot identifies residential estates with the highest proportion of teenage females and males aged 15 to 19, the demographic groups shown earlier to be most vulnerable to depression. Unlike Micro and Macro plots that establish patterns and underlying causes, this plot directly answers the question of where interventions should be implemented to maximize impact.

By highlighting estates with a higher concentration of teenagers at-risk of depression, the visualization provides policymakers with a data-driven basis to prioritize resources, such as school-based mental health programs, community counselling services and physical activity initiatives. This enables the government and relevant agencies to deploy targeted mental health policies at the estate level, ensuring that limited resources are allocated efficiently and interventions reach the populations most in need, thereby improving the overall effectiveness of depression prevention efforts in Singapore.



Plot 3: Percentage Of 15-19 Year Old Residents Across Singapore Estates (By Gender)

This visualization is presented using two subplots, with one map representing females aged 15 to 19 and the other representing males aged 15 to 19 across Singapore's planning areas. A color-blind friendly sequential color scheme is applied consistently across both maps. Darker shading indicates a higher proportion of teenagers while lighter shades represent lower proportions. Distinct hues (red for females and blue for males) are used to differentiate gender, while maintaining identical scales to allow direct visual comparison between the two subplots. This design ensures clarity, accessibility and ease of interpretation for policymakers.

Note that some areas of the Singapore map are in white and have no values. This is expected as the areas in white are estates with no residents – they are either industrial estates (eg. Boon Lay, Pioneer, Sungei Kadut), forested areas (eg. Central Water Catchment, Western Water Catchment, Mandai) or islands (eg. Sentosa, Pulau Ubin, Jurong Island).

Analysis And Observations:

From the visualization:

1) For **Teenage Females**, top five estates with the highest population proportions are:
    - **Pasir Ris** (4.29%)
    - **Choa Chu Kang** (4.10%)
    - **Woodlands** (4.08%)
    - **Bukit Timah** (3.42%)
    - **Tampines** (3.36%)

2) For **Teenage Males**, top five estates with the highest population proportions are:
    - **Pasir Ris** (4.52%)
    - **Choa Chu Kang** (4.40%)
    - **Woodlands** (4.33%)
    - **Tampines** (3.60%)
    - **Bukit Panjang** (3.45%)

3) These findings suggest that Pasir Ris, Choa Chu Kang and Woodlands consistently exhibit the highest concentrations of teenagers, regardless of gender. Consequently, these estates represent priority locations for the implementation of depression-reduction policies targeted at adolescents.

The substantial overlap between the two maps strengthens the case for estate-level, gender inclusive interventions, rather than highly segregated gender-specific deployment strategies. The near-identical ranking of estates for both teenage males and females indicates that teenage mental health challenges are spatially concentrated rather than gender-segmented. This implies that possible interventions can be implemented holistically within these estates, maximizing reach and cost-effectiveness.

Possible Reasons Top Estates Are Largely The Same Across Both Genders:

- **School Catchment And Accessibility:** Estates such as Tampines, Woodlands and Choa Chu Kang are well-served by secondary schools, junior colleges and polytechnics. This makes them attractive to families with school-growing children, as proximity to schools reduces commuting time and supports academic routines.

- **Affordability And Upgrading Pathways For Young Families:** These estates offer a relatively affordable entry point for young families, particularly through larger HDB flat types and upgrading pathways such as moving from smaller to larger flats within the same town. As families grow in size, they are more likely to remain within these estates rather than relocate, leading to the accumulation of teenage populations over time.

- **Lower Residential Turnover Compared To Central Areas:** Compared to central or prime districts (eg. Tanglin, Newton, Downtown Core, Outram), these estates exhibit lower residential turnover. Families are less likely to move frequently due to stable housing costs and established social networks. This residential stability allows cohorts of children to age in place, leading to noticeable clustering of teenagers.

Possible Solution Implementation To Reduce Depression Rates In Singapore:

Having identified the demographic groups most affected by depression, particularly teenage females aged 15 to 19, the key contributing factors (social support deficits, insufficient physical activity, academic or work-related stress and family history) and the Singapore estates with the highest concentrations of at-risk teenagers (Pasir Ris, Choa Chu Kang, Woodlands), the analysis now shifts towards translating insights into actionable policy strategies. These strategies aim to directly address the underlying causes of depression within the targeted estates, ensuring that interventions are both evidence-based and resource efficient.

The following interventions are designed to be estate-level, scalable and aligned with the identified risk factors of depression:

1. **Low Social Support Levels**
    - **Community-Based Youth Support Networks**
        - Establish youth mental wellness hubs within community centers and schools
        - Expand access to peer-support groups, counselling services and mentorship programs facilitated by trained professionals
        - Encourage structured after-school and weekend social activities to foster peer connection and emotional support

2. **Lack Of Physical Activity**
    - **Estate-Level Active Lifestyle Programs**
        - Introduce free or subsidized sports and fitness programs for teenagers at community sports facilities
        - Partner with schools and town councils to promote regular physical activity through inter-school or inter-estate sports leagues
        - Improve accessibility to parks, cycling paths and recreational spaces within estates

3. **Work Stress Levels**
    - **Stress Management And Academic Support Initiatives**
        - Implement stress management workshops focusing on coping skills, time management and resilience building
        - Provide school-linked academic support, such as study coaching and exam-preparation programs to reduce performance-related stress
        - Promote awareness of healthy work-life balance among students, parents and educators

4. **Family History/Home Environment**
    - **Family-Centered Mental Health Support**
        - Offer family counselling and parental education programs to help caregivers recognize early signs of depression and provide effective emotional support
        - Strengthen collaboration between schools, healthcare providers and family service centers to support at-risk households

----

Policy Implementation And Evaluation Strategy:

In order to ensure effectiveness and sustainability, the proposed interventions should follow a phased, evidence-driven implementation approach:

Phase 1: Pilot Implementation

- Launch pilot programs in the estates that have highest proportions of 15 to 19 year-old females and males (Pasir Ris, Choa Chu Kang, Woodlands), where the concentration of at-risk teenagers is the highest

- Coordinate efforts across schools, community centers, healthcare providers and social service agencies

Phase 2: Monitoring And Evaluation

- Track outcomes using standardized mental health indicators such as changes in reported depression levels, service utilization rates and school well-being surveys
- Conduct pre-intervention and post-intervention assessments to measure the changes and impact of policies on depression levels

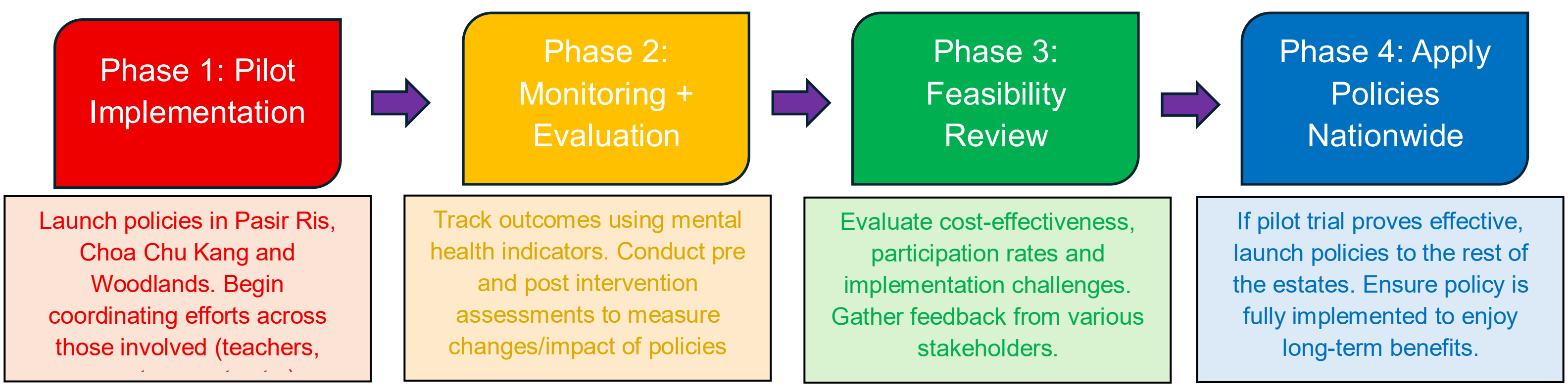
Phase 3: Feasibility And Effectiveness Review

- Evaluate cost-effectiveness, participation rates, implementation challenges and difficulties faced.
- Gather feedback from students, families, educators and service providers to refine programme design if there is a need to

Phase 4: Scaling Up Nationwide

- If the pilot demonstrates positive outcomes, progressively expand successful interventions to other estates across Singapore, prioritizing areas with similar demographic profiles
- Integrate proven strategies into existing national mental health frameworks to ensure long-term sustainability

Pictorial Representation Of Phases:



Limitations Of The Three Datasets Used:

While the analysis provides valuable insights into depression patterns and potential intervention strategies, several limitations associated with the datasets used should be acknowledged. These limitations may affect the precision, generalizability and interpretability of the findings.

Limitations Of Dataset 1: Depression DALY Contribution By Age Group And Gender

1) DALY As A Proxy For Depression Prevalence

Although the percentage contribution of DALYs attributed to depression serves as a useful indicator of disease burden across age groups and genders, it does not provide a perfectly accurate measure of depression prevalence. In particular, teenagers may experience multiple co-occurring mental health conditions such as anxiety disorders, obsessive-compulsive disorder (OCD) or other psychological conditions. As DALYs aggregate the burden across various conditions, the relative contribution of depression may be influenced by the presence of these comorbidities, potentially understating or overstating the true severity of depression among adolescents.

2) Limited Applicability To The Singapore Context

This dataset reflects global trends and is likely derived primarily from Western populations. Cultural, societal and healthcare system differences may result in depression patterns in Singapore that differ from those observed internationally. For instance, depression prevalence among older adults in Singapore may be higher or lower than global averages. For the purpose of the study, it is assumed that Singapore follows a broadly similar age and gender based trend, though this assumption may not fully hold in reality.

Limitations Of Dataset 2: Anxiety And Depression Mental Health Factors

1) Small Sample Size For Target Age Group

A key limitation of this dataset is the restricted sample size for the age group of interest. As the dataset includes respondents aged 18 and above and the analysis focuses on teenagers aged 15 to 19, only individuals aged 18 and 19 are available for analysis. This results in a relatively small number of observations (approximately 20 to 30 per gender), which increases statistical uncertainty and may reduce the reliability of the estimated correlation coefficients.

2) Subjectivity And Self-Reported Measures

The dataset is based on self-reported survey responses, which introduces subjectivity and measurement variability. Individuals may interpret and rate mental health indicators differently. For example, two respondents with similar depression states and levels may assign substantially different scores. This subjectivity extends to related factors such as anxiety levels, financial stress and work stress, potentially introducing noise and bias into the correlation analysis.

3) Incomplete Coverage Of Depression Risk Factors

The dataset includes only a limited set of potential contributing factors to depression. While variables such as social support and physical activity are informative, the list is not exhaustive. Important real-world factors such as parental abuse, social media usage, cyberbullying and traumatic life events are not captured in the dataset. Consequently, the analysis can only identify the most significant factors within the scope of the dataset. The findings should not be interpreted as definitive evidence of the primary causes of teenage depression.

Limitations Of Dataset 3: Singapore Resident Population By Planning Area/Subzone, Age Group And Sex (General Household Survey 2015)

1) Outdated Demographic Information

The dataset is based on statistics published in 2015, making it over a decade old. Given the rapid pace of urban development and demographic shifts in Singapore, current population distributions may differ substantially. Estates such as Pasir Ris, Woodlands and Choa Chu Kang may no longer have the highest proportions of 15 to 19 year old males and females currently. Furthermore, newer residential towns such as Punggol, Sengkang and Tengah may be underrepresented or absent, leading to potential misidentification of priority intervention areas.

2) Use Of Rounded And Estimated Values

The population figures in this dataset are estimated and rounded, as evidenced by many values ending in zeros. While suitable for high-level analysis, this rounding may obscure small but meaningful differences between estates. Minor changes in population estimates could alter the calculated proportions of teenagers aged 15 to 19, potentially affecting estate rankings in the actionable analysis.

Taken together, these limitations in the three datasets highlight the need to interpret the results with caution. While the datasets used provide a useful foundation for exploratory analysis and policy discussion, future research would benefit from Singapore specific, up-to-date and larger scale datasets with more granular mental health indicators. Such improvements would strengthen the accuracy and policy relevance of subsequent analysis.

Conclusion:

This project demonstrates how data visualization can move beyond description to inform targeted mental health interventions. By analyzing depression rates grouped by gender and age group, we identify teenage females aged 15 to 19 as a particularly vulnerable group and highlight specific estates in Singapore where interventions may have the greatest impact. Importantly, the findings emphasize that depression is shaped by multiple, interrelated factors such as Social Support and Physical Activity, rather than a single cause. While the analysis is constrained by the limitations of self-reported and observational data, it provides a defensible, evidence-based starting point for piloting place-based mental health strategies. Future work can build on this approach by incorporating up-to-date data and evaluating the effectiveness of implemented interventions, ultimately strengthening the translation of data insights into public health action.