Data topic area: healthcare

Data set: <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease> (specifically the 2022 dataset)

Topic: We will train an algorithm to recognize heart disease symptoms and predict if a patient is at risk.

Research questions (as time and restrictions allow) :

1. Are males or females more likely to be at risk for heart disease?
2. Is age a factor in heart disease risk?
3. Is there a relationship between physical activity and heart disease?
4. Are there any variations in regions (states) of the dataset in terms of risk of heart disease?
5. Is race a factor in heart disease risk? For example, do white people experience higher instances of heart disease?
6. Does smoking have an impact on your risk of heart disease?
7. Does drinking alcohol have an impact on your risk for heart disease?
8. Is BMI an indicator of heart disease risk?
9. Is asthma an indicator of heart disease risk?
10. Is kidney disease an indicator of heart disease risk?
11. Is having a stroke previously an indicator of future heart disease?
12. If you have difficulty with walking, does that indicate heart disease?
13. Is having diabetes an indication of heart disease?
14. If your general health is fair or poor, is that an indicator for heart disease?
15. Does the amount of sleep that you receive on average pose a risk for heart disease?
16. Does being deaf or blind have an impact on heart disease?

Tasks:

1. Data cleaning (Larry)
2. Visuals - heat map (Carlie), graphs and other visuals (Carlie and Emely)
3. Machine learning (Larry, Carlie, Emely)
4. Presentation (Emely)
5. ReadME (Emely, with checking by Larry and Carlie)