Donna Mulkern
COMP4447 Final Project
Dementia and Alzheimer's: Early Predictors

Github: https://github.com/donnamulkern/COMP4441-FinalProject

DATASET
The data used in this project comes from the OASIS (Open Access Series of Imaging Studies) Brains organization, which aims to provide neuroscientific images and data freely to the scientific community. The grant citations for both datasets are below:

OASIS: Cross-Sectional: Principal Investigators: D. Marcus, R, Buckner, J, Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382

OASIS: Longitudinal: Principal Investigators: D. Marcus, R, Buckner, J. Csernansky, J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382

This dataset was accessed through Kaggle at https://www.kaggle.com/jboysen/mri-and-alzheimers and directly from the OASIS website: https://www.oasis-brains.org/#access

MOTIVATION
Dementia and its specific subtype of Alzheimer's disease are some of the most pervasive health conditions that affect millions of people around the world. A diagnosis of this disease is heartbreaking not only for the patient but also for their families, because this means that over time, the patient will slowly lose touch with the world around them and might even stop recognizing their loved ones.

Brain scans of Alzheimer's and dementia patients indicate loss of white matter volume, deteriorated neuronal connections, and a buildup of amyloid plaques. Taken all together, the patients exhibit symptoms such as short-term memory loss, general forgetfulness, increasing dependence on caregivers, speech loss, and other coordination problems.

Although a drug called Aduhelm was recently approved to be fast-tracked by the FDA as a treatment for Alzheimer's disease (https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-approval-alzheimers-drug), the evidence for the effectiveness of the drug is not overwhelmingly positive, and at best, the drug will only slow down the deterioration.

Given the widespread nature of the disease and the burden that it places on society as a whole, studying this disease and identifying signs early on is of utmost importance. The focus of this project will be to attempt to identify the correlation of dementia onset with other factors.

RESEARCH QUESTION

This project will be examining two separate datasets-- one that is cross-sectional (a larger group at a single moment in time) and one that is longitudinal (data collected from a smaller group over a longer period of time). The data use agreement restricts anyone from attempting to combine these datasets due to public health concerns, but even so we can investigate a few questions from different angles using these two datasets.

Can we predict a diagnosis of Alzheimer's disease or dementia from other demographic factors? Is there any correlation in how fast the symptoms accelerate and other demographic factors? Can we predict the likelihood of a patient developing dementia or Alzheimer's disease from demographic information?

LITERATURE REVIEW

Ardekani, Babak A et al. "Corpus callosum shape changes in early Alzheimer's disease: an MRI study using the OASIS brain database." Brain structure & function vol. 219,1 (2014): 343-52. doi:10.1007/s00429-013-0503-0

The above article used brain scans from the OASIS brain database to visualize differences in the corpus callosum. The corpus callosum is the connection between the two hemispheres of the brain; decay of this bundle of fibers results in decreased cognitive coherence. The study found that decreased circularity of the corpus callosum is correlated with early Alzheimer's disease. There is also sexual dimorphism associated with corpus callosum circularity; the corpus callosum is less circular in female brains than in male brains.

Chen, Ting et al. "CAVIAR: CLASSIFICATION VIA AGGREGATED REGRESSION AND ITS APPLICATION IN CLASSIFYING OASIS BRAIN DATABASE." Proceedings. IEEE International Symposium on Biomedical Imaging vol. 2010 (2010): 1337-1340. doi:10.1109/ISBI.2010.5490244

The above article also utilized the OASIS brain database images, but took a different approach in that they created the CAVIAR algorithm for classification from a training set of images, and then validated it using a different set of images. The algorithm is based on "the histogram of the deformation field between the MRI brain scan and the atlas which captures the structural changes in the scan with respect to the atlas brain". The classifier was shown to be a good way of identifying different brain states of healthy or diseased and what stage of dementia was present.

The current project will be smaller in scope because I simply do not have the computing power required to view the MRI scans themselves and analyze that data; however, I will be able to view the external subject demographic information and draw conclusions based on that.

CONCLUSIONS:

I used normalized Whole Brain Volume (nWBV) to approximate dementia in analysis to avoid having to regress against a categorical variable. Using the cross-sectional dataset, I found that using least squares regression and all demographic and brain measurement variables, the model has an R squared value of .664, which is a promising value for regression. For further analysis, I would recommend classification instead of regression so that the categorical variable could be utilized and predictions regarding dementia diagnosis can be made. The correlations are found in *Figure 1*.

The correlation heat map for the longitudinal dataset is shown in *Figure 2*. This dataset gave a glimpse into the differences between demented and non demented subjects, as well as those who were diagnosed throughout the timeframe of the study. The normalized Whole Brain Volume variable was immensely affected throughout time and was different for each subject each MRI visit, which is unsurprising because Age is one of the strongest predictors of this variable. This illustration is shown in *Figure 3*.

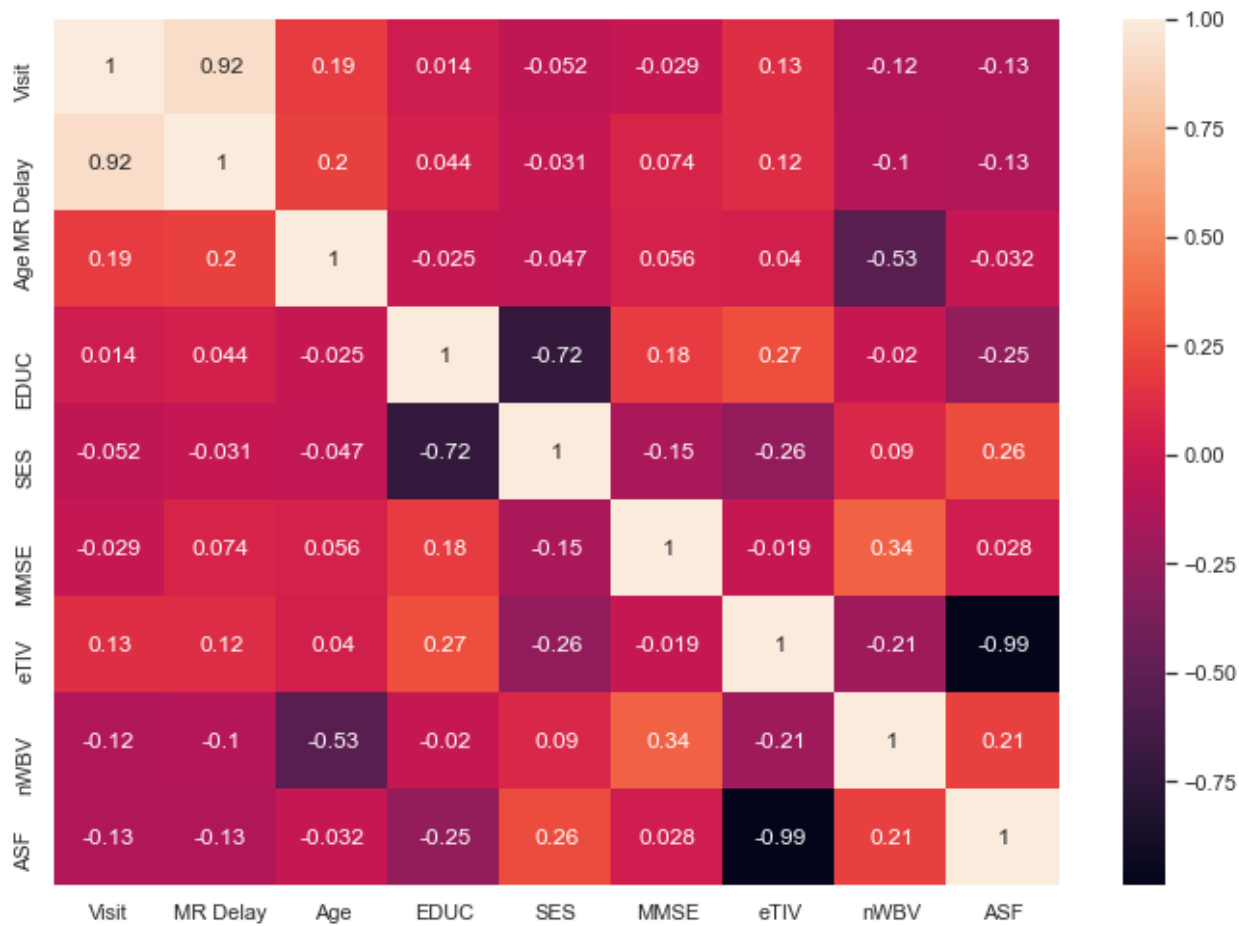**APPENDIX: VISUALIZATIONS**

Figure 1

Figure 2

Figure 3