WESTERN MICHIGAN UNIVERSITY

CS- 5610
Advanced R for Data Science

<u>Project</u>

<u>EXPLORATORY DATA ANALYSIS FOR PREOPERATIVE BRAIN GLIOMA MRI</u>

Submitted by:

Swapnil Rajendra Patil   **Varshitha Chava**
Win ID: 453347613   **Win ID: 390503567**

Submitted to:
Wassnaa AL-mawee

Motivation and Overview:

The brain is an important organ in the human body. But sometimes organs are affected by various diseases that affected human bodies. The brain is also get affected by various diseases that may lead a human to death. So, it is important to know about the diseases that can cause severe brain issues. Here we took an approach where we will deal with various information that is actually there to create brain disease. In other words, we can say that here we will learn about various factors that actually cause brain disease and also make an exploratory data analysis to understand these disease-related factors that can cause brain problems in the human body.

Related work:

Like the other domains, data researchers are also making their impact in medical domains as well. A variety of diseases and their related information is hereby analyzed to understand the disease better and thus it can also conclude the precautionary measures that one can take to get rid of various diseases. In this scope of work, we will discuss brain disease… Al has developed a work to make a prediction of brain disease and they also analyze the various factors that are correlated with this disease in their research. Al also makes research on the exploratory data analysis of brain disease to understand the correlated factors and along with that the prediction of brain disease on the basis of these correlated attributes.

Initial questions:

The proposed work is developed to analyze and understand the set of information first. Then based on the reports we will here develop a Shiny App application where we will analyze the brain diseases related factors and also make a plot to understand the correlated attributes that are present in the proposed brain disease dataset.

Data:

A cancer imaging archive is used for the collection of data for the analysis to be made. Includes a population consisting of 501* adult patients. Confirmed grade II-IV diffuse gliomas who underwent preoperative MRI, tumor resection, and tumor genetic testing at a single medical Center**.

Term	Represents	Values
ID	DICOM (0010,0020) PatientID	
Sex	DICOM (0010,0040) Patient Sex	M,F
Age at MRI	Age in years at time of MR imaging	
WHO CNS Grade	Grade per the 2021 World Health Organization Classification of Tumors of the Central Nervous System (WHO CNS 2021) (https://doi.org/10.1093/neuonc/noab106)	2,3,4

Term	Represents	Values
Final pathologic diagnosis (WHO 2021)	Final (integrated) pathologic diagnosis per the 2021 World Health Organization Classification of Tumors of the Central Nervous System (WHO CNS 2021) (https://doi.org/10.1093/neuonc/noab106)	(IDH) -wildtype, - IDH-mutant, - Astrocytoma, IDH-wildtype, - Oligodendroglioma, IDH-mutant, 1p/19q-codeleted
MGMT status	O6-methylguanine-DNA methyltransferase status - clinical interpretation of the MGMT index described below.	negative, positive, indeterminate
MGMT index	O6-methylguanine-DNA methyltransferase methylation index (in house method developed by UCSF clinical labs, https://genomics.ucsf.edu/content/mgmt-promoter-methylation-assay) where numeric values 0-17 indicate the number of promoter methylation sites.	0-17, blank
1p/19q	presence of codeletion of 1p and 19q genes, assayed by fluorescent in-situ hybridization	intact, co-deletion, relative co-deletion, unknown
IDH	isocitrate dehydrogenase mutation subtype characterized with a capture-based targeted next-generation DNA sequencing panel (UCSF500) as described in (https://doi.org/10.1093/neuonc/now254)	
1-dead 0-alive	Survival status of the patient at last clinical follow up.	
OS	Overall survival in days from initial diagnosis to last clinical follow up.	
EOR	Extent of resection determined by review of operative reports and immediate postoperative imaging	biopsy (only biopsy) Subtotal resection (STR) gross total resection (GTR)
Biopsy prior to imaging	Was a burr-hole biopsy performed prior to imaging	yes, no, blank

Exploratory data analysis:

As we have mentioned earlier that we will create a Shiny App using a brain disease dataset in the R programming environment. So here in the user interface, we selected some information from the user choice and based on the selected choices, we made an analysis, plotted histogram and tabular data to understand the entire dataset.

For predictions: As we provide the inputs for the random scenario it will shiny app will predict the percentage of survival of the patient.

Positive outcome is below 0.5 which says patient may survive with the treatments.

Negative outcome is above 0.51 where patients may not survive.

May survive

May not Survive

For EDA: This bar graph is obtained by variable Age at MRI which shows that most of the patients suffering from brain glioma are between 51 to 78 years of age.

Visualization:
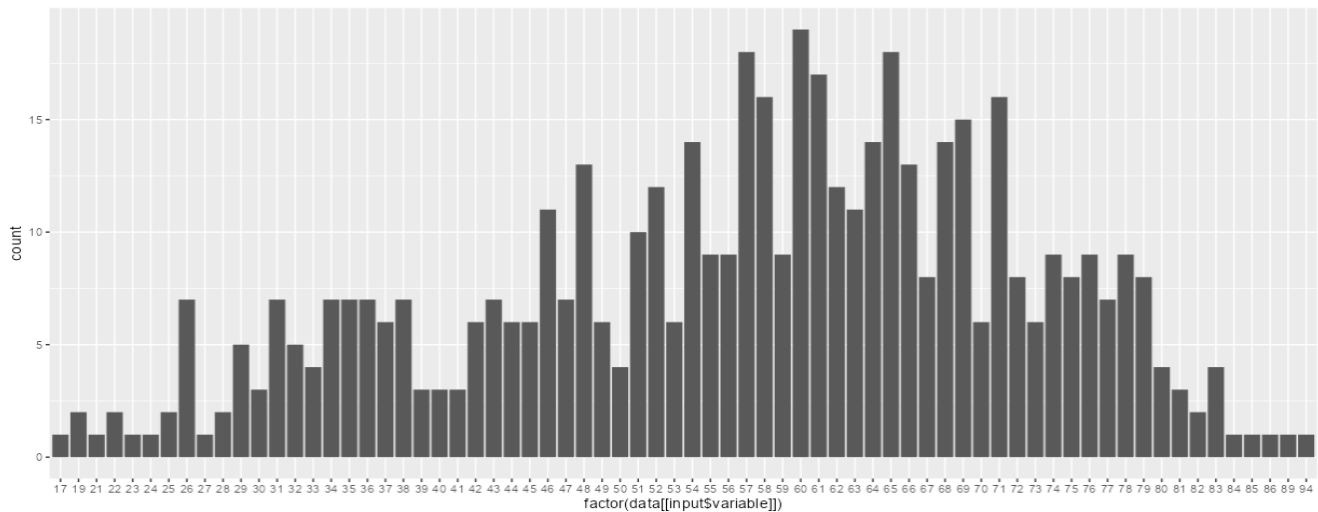
Below is second example from the analysis with final pathological diagnosis which shows that more than 350/501 patients are suffering with Glioblastoma(wildtype) which is most deadliest brain glioma.

Visualization:
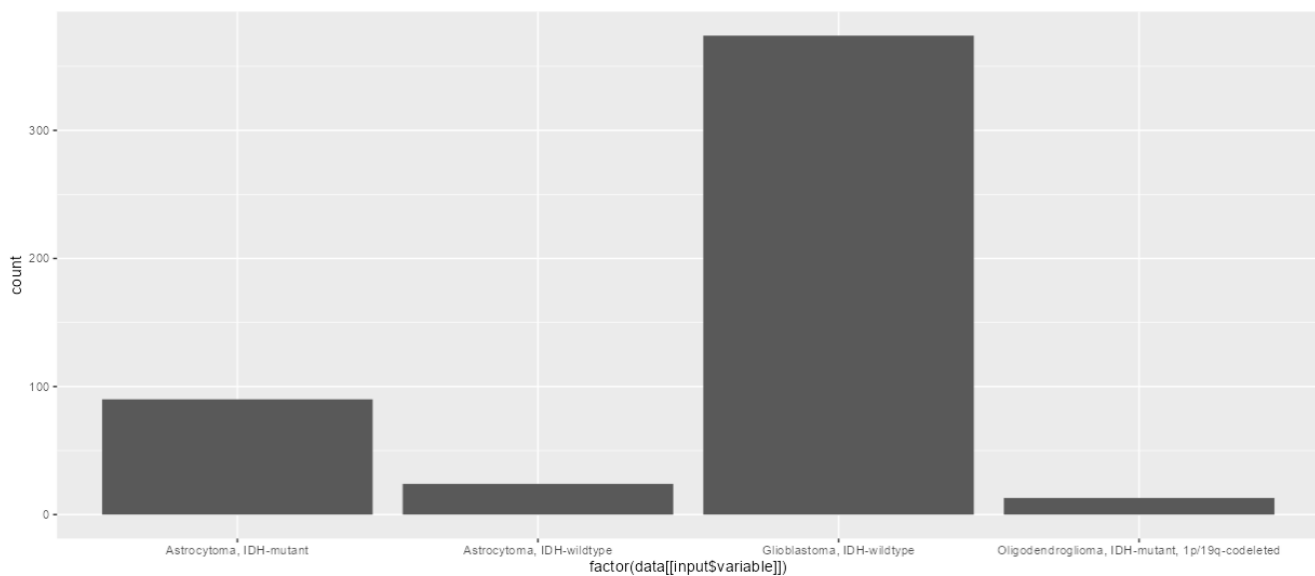
Data analysis:

There is multiple information that can be used to define exploratory data analysis of the brain glioma disease dataset. We may need multiple library packages. As soon as we read the dataset we jumped into the plotting of the information and blend them with their correlated features. For this, we used scatterplots, bar plots, box plots etc. for the analysis.

Narrative and summary:

For the EDA part, the dataset was explored using different graphical and statistical techniques such as scatterplots, bar plots, and box plots. The analysis focused on different variables, such as Age at MRI and final pathological diagnosis. It was found that most patients suffering from brain glioma were between 51 to 78 years of age, and more than 350 out of 501 patients were suffering from the deadliest type of brain glioma, Glioblastoma (wildtype).

For the prediction part, a Shiny App was created to predict the percentage of survival of the patient based on user input. It was mentioned that there can be 133,056 different predictions for patients with their conditions of disease.

Overall, the project provides a detailed analysis of brain glioma disease using a dataset and different analytical techniques in R. The Shiny App also provides a predictive model to help clinicians and patients make informed decisions.