**Übersicht über die eigene Beteiligung :**

Am Anfang habe ich die Aufgaben im Kontext der Anforderung in einer Word Datei (Tasks im Git-Repo) zusammengefasst .Diese Datei versuchte ausführlich wie möglich zu formulieren ,damit jeder seinen Aufgabenbereich aussuchen kann. Drin enthält zu jedem Stichpunkt(Tabs , Graphen ,Tabllen ,Anpassung des Datensatz )alle Schritten ,die im Rahmen der Appimpementierung benötigt werden z.B: Anpassungschritten(Fehlende Werte umgehen , welche neue Variable sollen angefügt ?, encoding …. ), Graphen Typen (welche Variable (diskret,stetig) mit welcher Variable eine Visaulisierung sinn macht ) .Auch habe ich die erwartete Schlüsse bei den Teststatistiken gefasst , was wir zu jeder Tabelle testen wollen (Unabhängigkeit ,..).Danach half ich die andere Mitglieder beim Clonen des Git-Repo ,sowie bei den Hauptfunktionen (commit , push..) anzuleiten . Mein Aufgabebereich war die zusammenfassende Tabellen zu erstellen, mit geeigneten Teststatistiken zu kombonieren , letztendlich die Ergebnisse zu interpretieren.Daneben ist die passende Maßen anzuwenden ,um sinnvolle Schlüsse aus den Tabellen zu ziehen . Außerdem unterstützte ich die Mitglieder bei der Anpassung des Datensatz und bei der Erstellung der Graphen.Zusätzlich zu den Problemlösungen zu finden ,war berücksichtigt alle Anforderung fehlerlos zu implementieren . Ich versuchte immer erste Schritten mit Beispielen zu schaffen , um die andere identische Schritte zu verfolgen, und erfolgreich mit anderen Stücken zusammenzusetzen .Am Ende fügte ich die Interaktionen an den Graphen an ,wobei unsere App breite Visaulisierungsmöglichkeiten anbieten kann ,z.B die Befüllung der Säulen beim Graph 3 nach einer durch user gewählten diskreten Variable mit den Prozentanteile von allen Stichproben durchführen .

**Fazit**

Die geplannte Aufgaben waren möglich fehlerlos zuimplementieren , da ich möglichst jeder Stichpunkt zu machbaren Unterpunkten eingeteilt habe , um die Herausforderung bezüglich der unausreichenden praktischen Kenntnissen zu beseitigen , früh zu erkennen .Den Schwergrad der Projekt würde ich auf 4 von 10 setzen , da ich genügend Hilfe von den Vorlesungsfolien bekommen habe . Auch der Coding-Anteil macht mir immer Spaß ,und die Problemlösung moteviert mich sehr . Durch dieses Projekt vertiefte ich meine Kenntnisse in R und kennte ich die breite Nutzungsmöglichkeiten der Shiny- Dashboards .Von diesem Projekt nehme ich die vielfältige Erfahrung bei der Problemlösung in neuer Sprache R mit . Der gemeinsamer Erfolg beim Einhalten des Deadline und beim Aufbau unserer App moteviert mich viel , um in Zukunft weitere Tools in R zu lernen .Aber ich glaube , als ob der Datensatz mehrere Spaltengehabt hätte , könnten wir noch mehr Analysentechniken umsetzen ,die wir im Rahmen der Statistikmodulen gelernt haben .

**Umsetzung der App**

Danks der Tasks Datei konnten wir uns vorstellen , wie die App aussehen wird , und welche Bibliotheken im Rahmen der Umsetzung benötigt werden .

Die App wurde mit Hilfe von Shiny Bibliothek aufgebaut , deswegen sollten wir zwei Skripte erstellen ,einer für UI ,anderer für Server der App .

Unsere App besteht aus 4 Tabs in einem tabsetPanel,das auf einem mainPanel vom sidebarLayout dargestellt ist :

* Data table : auf diesem Tab wird der angepasste Datensatz gezeigt .wie erwähnt , wurden die Anpassungsschritten wie geplannt implementiert . Diese Tabelle wurde als ein Instanz vom Typ renderTable im output zurückgegeben ,Dann wird es im tableOutput auf dem User Interface gelesen und ausgegeben.
* Graphen : besteht aus 3 verschachtelten Tabs :

1. Scatter Plot : Das Bibliothek ggplot2 bietet uns mit der Funktion geom\_point zwei stetige Variablen zu visaulisieren an . Auf dem wird die Korrelationstärke/-richtung dazwischen mit Hilfe von Funktion geom\_smooth dargestellt . Auswahl der Variablen : Durch selectInput wird eine Selectlist angezeigt , enthält als Auswahlen drei stetige Variablen :Cholesterinwerte ,Tumorgrößen ,BMI(Body Mass index).Der ineraktive Knopf („update“) erstellt den entsprechenden Graph mit der Möglichkeit durch Radiobuttons eine diskrete Variable auszusuchen , um die Punkte im Scatterplot nach dieser Variabke zu färben .
2. Box Plot : In demselben Bibiliothek besteht eine Funktion unter dem Name geom\_boxplot ,hier macht es sinn zu schauen ,ob ein Unterschied in Tumorgrößen bei den Gruppen .Die Einteilung der Patienten in den bestimmten Gruppen ,wird entweder auf Rauchenverhältnis oder auf Geschlecht basiert . Hierfür habe ich Radiobuttons zum Aussuchen angewendet.Zusätzlich gibt es Möglichkeit ohne Einteilung mit None es darzustellen .
3. Säulendiagram : Auf X-Achse werden die Altersgruppen versetzt .Mein Ziel war hier die relative und absolute Häufigkeit eine diskrete Variable zu schauen .Auf Y-Achse werden die absolute Häufigkeiten darzustellen ,und die Säulenbefüllung wird durch diese diskrete zu zeichnen . Der Auswahl erfolgt auch durch radioButtons Komponent .Die Variable sind Tumorgrößendichotom ,Cholesterindichotom und Geschlecht . Die Zeichnung eines Säulendiagrams wird mit Hilfe von Funktion geom\_bar durchgeführt . Ein Beispiel : Im Fall nach Cholesterindichotom die Säulen zu befüllen ,sehen wir die Altersgruppe zwischen45-59 , auf dieser Säule wurde im blauen Teil 29.7% , das bedeutet : 29.7% aller Patienten von allen Altersgruppen , sind im Alterbereich von 44-59 Jahre alt **und** haben normale Cholesterinwerte(Schnittsmenge) .

Alle Graphen wurden als Instatnz von der Klasse renderPlot im output

zurückgegeben ,und werden auf User Interface mit Hilfe von plotOutput gezeichnet .

* Tabellen : besteht aus 3 verschachtelten Tabs :

1. Erste Tabelle ist eine Kreuztabelle zwischen die nach TNM klassifizierte Tumorgrößen und Altersgruppen , mit Hilfe von der Interngebauten Funktion xtabs . um die Tabelle richtig zu formulieren ,sollte die Tabelle im Form vom Datentyp data.frame.matrix ,dann im Instanz von der Klasse renderTable zurückgegeben wird . Auf der User Oberfläche im tableOutput wird dieses Instanz angezeigt .

|  | **30-44** | **45-59** | **above 60** | **under 30** |
| --- | --- | --- | --- | --- |
| T1 | 7 | 0 | 0 | 2 |
| T2 | 77 | 165 | 49 | 0 |

Anschaulich gibt es keine Patient über 45 Lebensjahr , deren Tumoren T1(<2 cm) nach TNM klassifiziert .Sowie besteht es keine junge Patienten an T2 Tumoren betroffen .

Um die Abhängigkeit zwischen Alter und Tumorgrößen wird mit einem interaktiven Button(„Abhängigkeit testen“) komboniert . Teststatistik : ich konnte der Chi-Quadrad Test nicht anwenden , weil die mehr als 20% aller Zellen der Tabelle sind kleiner als 5. Deswegen sollte Exakter fisher Test durchgeführt werden .

P-Wert ist viel kleiner als 5% , Der Zusammenhan zwischen Alter und Tumorgrößen kann nachgewiesen werden .

1. Zweite Tabelle ist eine Kreuztabelle zwischen Rauchen und Tumorgrößen ,identische Schritte verfolgt .

|  | **T1** | **T2** |
| --- | --- | --- |
| nicht-raucher | 3 | 152 |
| raucher | 6 | 139 |

Ein sinnvolles Maß angewendet wurde ,ist Odds Ratio ,da es uns für den Vergleich der Risiken bei beiden Gruppen interessieren .Im Bibliothek epiR finde ich die Funktion epi.2by2 ,um das Odds Ratio und Konfidenzintervall zu bestimmen . ODDS Ratio beträgt 0.457 ,das bedeutet ,das Risiko bei Raucheren an großen Tumoren zu erkranken ist um 54 % höher als bei den Nicht-Raucheren .

1. Dritte Tabelle ist eine Kreuztabelle zwischen Cholesterinwerte und Tumorgrößen ,

|  | **hoch** | **niedrig** | **normal** |
| --- | --- | --- | --- |
| T1 | 0 | 8 | 1 |
| T2 | 58 | 143 | 90 |

Aus der Tabelle können wir sehen , dass es keine Patienten gibt ,die höhe Cholesterinwerte und große Tumoren haben . Hier ist festzustellen , dass die Tumorgrößen bei allen Cholesterin-Gruppen gleich ist .

Hier wollte ich untersuchen ,ob die Cholesterinwerte bei beiden Tumorklassifikationen gleich ist . Deswegen sollte ich mit dem Formel von erwarteten Häufigkeiten eingesetzt ,um die Erfüllung der Voraussetung vom Homogenitätstest zu bestätigen .

|  | **hoch** | **niedrig** | **normal** |
| --- | --- | --- | --- |
| T1 | 1.74 | 4.53 | 2.73 |
| T2 | 56.26 | 146.47 | 88.27 |

P-Wert beträgt 0.058 ist größer als 0.05 % , d.h es spricht nichts gegene die Gleichheit der Tumorgrößen in allen Cholesteringruppen .

* About : Enthält paate allgemeine Informationen ,z.B warum wir Lungenkrebs ausgewählt haben ,Namen von Teammitgleider , welche Packages wir im Rahmen der Projektarbeit benutzt haben .