

```

library(tidyverse)
library(plyr)
library(dplyr)
library(ggplot2)
library(dplyr)
library(scales)
library(readr)
require(GGally)
library(caTools)
library(corrplot)
library(e1071)
library(caret)
library(klaR)
library(randomForest)
library(Hmisc)
library(ROCR)
library(pROC)
library(party)


data <- read_csv("C:/Users/Jaswanth/Desktop/DataManagement/Lab3
Classification/data_class.csv")
data_num <- read_csv("C:/Users/Jaswanth/Desktop/DataManagement/Lab3
Classification/data_num.csv")


data<- as.data.frame(data)
data_num<- as.data.frame(data_num)


view(data)
view(data_num)


#Region wise
# Hospitalized
ggplot(data, aes(x= Region,fill=Hospital_status)) +
geom_bar(position=position_dodge(width=0.9)) +
  geom_label(stat='count',aes(label=..count..),
size=4,position=position_dodge(width=0.9),vjust=-0.1)+
  labs(y="Count of Infected people",subtitle="Total number of Infected people
Hospitalization status")


#Death
ggplot(data, aes(x= Region,fill=Death)) + geom_bar(position="stack")+
  geom_label(stat='count',aes(label=..count..),
size=4,position="stack",vjust=-0.1)+
  labs(y="Count of people died",subtitle="Number of People died Region wise")


#Age group graph


#Showed Symptoms or not
ggplot(data, aes(x=Age_group,fill=Asymptomatic),group=Gender) +
geom_bar(position="stack")+

```

```

    geom_label(stat='count',aes(label=..count..),
size=4,position="stack",vjust=-0.1)+
    facet_wrap(~Gender)+labs(x="Age Group",
                                y="Count of Infected people",subtitle="Total number
of Infected people symptoms status")

```

```

#Died?
data_death=filter(data,Death=='Yes')
ggplot(data_death, aes(x=Age_group,fill=Asymptomatic),group=Gender) +
geom_bar(position="stack")+
    geom_label(stat='count',aes(label=..count..),
size=4,position="stack",vjust=-0.1)+
    facet_wrap(~Gender)+labs(x="Age Group",
                                y="Count of people died",subtitle="Total number of
people died with symptoms")

```

```

#Hospital
data_symp=filter(data,Asymptomatic=='Yes')
ggplot(data_symp, aes(x= Age_group,fill=Hospital_status),group=Gender) +
geom_bar(position=position_dodge(width=0.9))+
    geom_label(stat='count',aes(label=..count..),
size=4,position=position_dodge(width=0.9),vjust=-0.05)+
    facet_wrap(~Gender)+labs(x="Age Group",
                                y="Count of people",subtitle="Total number of
people hospitalized without showing symptoms")
#Died?

```

```

ggplot(data_symp, aes(x=Age_group,fill=Death),group=Gender) +
geom_bar(position="stack")+
    geom_label(stat='count',aes(label=..count..),
size=4,position="stack",vjust=-0.1)+
    facet_wrap(~Gender)+labs(x="Age Group",
                                y="Count of people died",subtitle="Total number of
people died without showing symptoms")

```

```

ggplot(data_symp, aes(x=Age_group,fill=Transmission),group=Gender) +
geom_bar(position="stack")+
    geom_label(stat='count',aes(label=..count..),
size=4,position="stack",vjust=-0.1)+
    facet_wrap(~Gender)+labs(x="Age Group",
                                y="Count of people Infected",subtitle="Total number
of people infected without showing symptoms")

```

```

#Correlation
corrplot(cor(data_num),method = "color")

```

```

#As factor

```

```

data_num$Death <- as.factor(data_num$Death)
#Train Test
set.seed(101)

```

```

split = sample.split(data_num$Death, SplitRatio = 0.70)
train = subset(data_num, split == TRUE)
test = subset(data_num, split == FALSE)

#SVM

svm_model <-svm(Death ~ Age_group + Asymptomatic+ Hospital_status,
               data = train,
               type = 'C-classification',
               kernel = 'linear')

summary(svm_model)

y_pred_svm = predict(svm_model, test)

confusionMatrix(y_pred_svm, test$Death)

#Logistic Regression

logisticregression <- train(Death ~ Age_group + Asymptomatic+ Hospital_status,
data=train, family=binomial,method='glm')

summary(logisticregression)
preds <- predict(logisticregression,test)
confusionMatrix(preds,as.factor(test$Death))

#Decision Tree with 3
Decision_tree <- ctree(Death~Age_group + Asymptomatic+ Hospital_status,
data=train)
plot(Decision_tree)
pred_dt <- predict(Decision_tree, test)

confusionMatrix(pred_dt, test$Death)

Decision_tree <- ctree(Death~., data=train)
plot(Decision_tree)
pred_dt <- predict(Decision_tree, test)

confusionMatrix(pred_dt, test$Death)

#Random Forest

Random_Forest <- train(Death~., data=train, method='rf')

Random_Forest <- train(Death~Age_group+Asymptomatic+Hospital_status,
data=train, method='rf')

Random_Forest$results

```

```
pred_rf <- predict(Random_Forest, test)
confusionMatrix(pred_rf, test$Death)

#Feature Importance
rf_imp <- varImp(Random_Forest, scale = FALSE)
rf_imp <- rf_imp$importance
rf_gini <- data.frame(Variables = row.names(rf_imp), MeanDecreaseGini =
rf_imp$Overall)

ggplot(rf_gini, aes(x=reorder(Variables, MeanDecreaseGini),
y=MeanDecreaseGini, fill=MeanDecreaseGini)) +
  geom_bar(stat='identity') + coord_flip() + theme(legend.position="none") +
labs(x="Variables", y="Importance") +
  ggtitle('Variable Importance Random Forest') + theme(plot.title =
element_text(hjust = 0.5))
```