```
# Insights on Student Results Dataset

# load library to read data from csv file

library('readr')

# import the dataset

Studentresult<- read_csv("C:/Users/MY SYSTEM/Desktop/Ireland Documents/SEMESTER
1/Visualization/Assignment 4/studentresult.csv")

# view the dataframe containing imported data

View(studentresult)

# check the structure of dataframe

str(studentresult)

# check number of rows of dataframe

nrow(studentresult)

# check number of columns of dataframe

ncol(studentresult)

# check length of dataframe

length(studentresult)

# get the names of each column

names(studentresult)

# get the dimensions of dataframe

dim(studentresult)

# get summary for every dimension in dataframe

summary(studentresult)

# get first 6 rows of dataframe

head(studentresult)

# get last 6 rows of dataframe

tail(studentresult)

# Visualization

# load libraries for plotting the data

install.packages('ggplot2')

library('ggplot2')

# load library to perform sql function on dataframe

install.packages('sqldf')
```

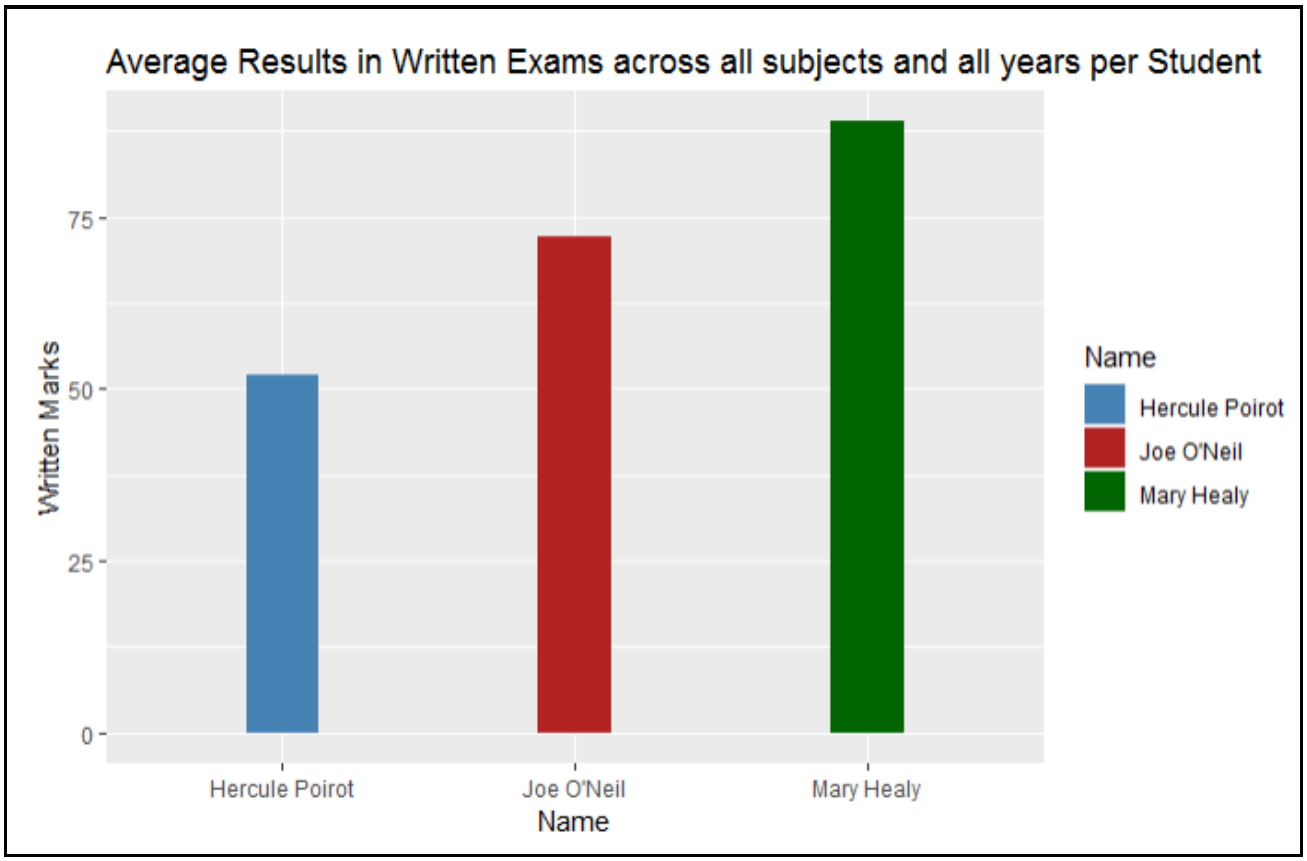
```
library('sqldf')

library('gsubfn')

library('proto')

library('RSQLite')
```

Average results in written exams across all subjects and all years per student

```
markswritten<-sqldf('select Name,avg(Mark_Written) as Written_marks from studentresult group by Name')

markswritten


# without sql

markswritten2<-aggregate(data=studentresult,Mark_Written~Name,FUN = mean)

markswritten2


# impute missing values

# calculate average score for oral marks and replace NA with that


# Version 1 with sqldf

avgmarks<-sqldf("select avg(Mark_Oral) from studentresult where Name='Mary Healy' and Mark_Oral is not 'NA'")

avgmarks


# Version 2 without sqldf

avgmarks2<-mean(studentresult$Mark_Oral[studentresult$Name=='Mary Healy' &
!is.na(studentresult$Mark_Oral)])

avgmarks2


# replace NAs with calculated average

studentresult$Mark_Oral<-
ifelse(is.na(studentresult$Mark_Oral),as.numeric(avgmarks),studentresult$Mark_Oral)

View(studentresult)


# graph student results
```

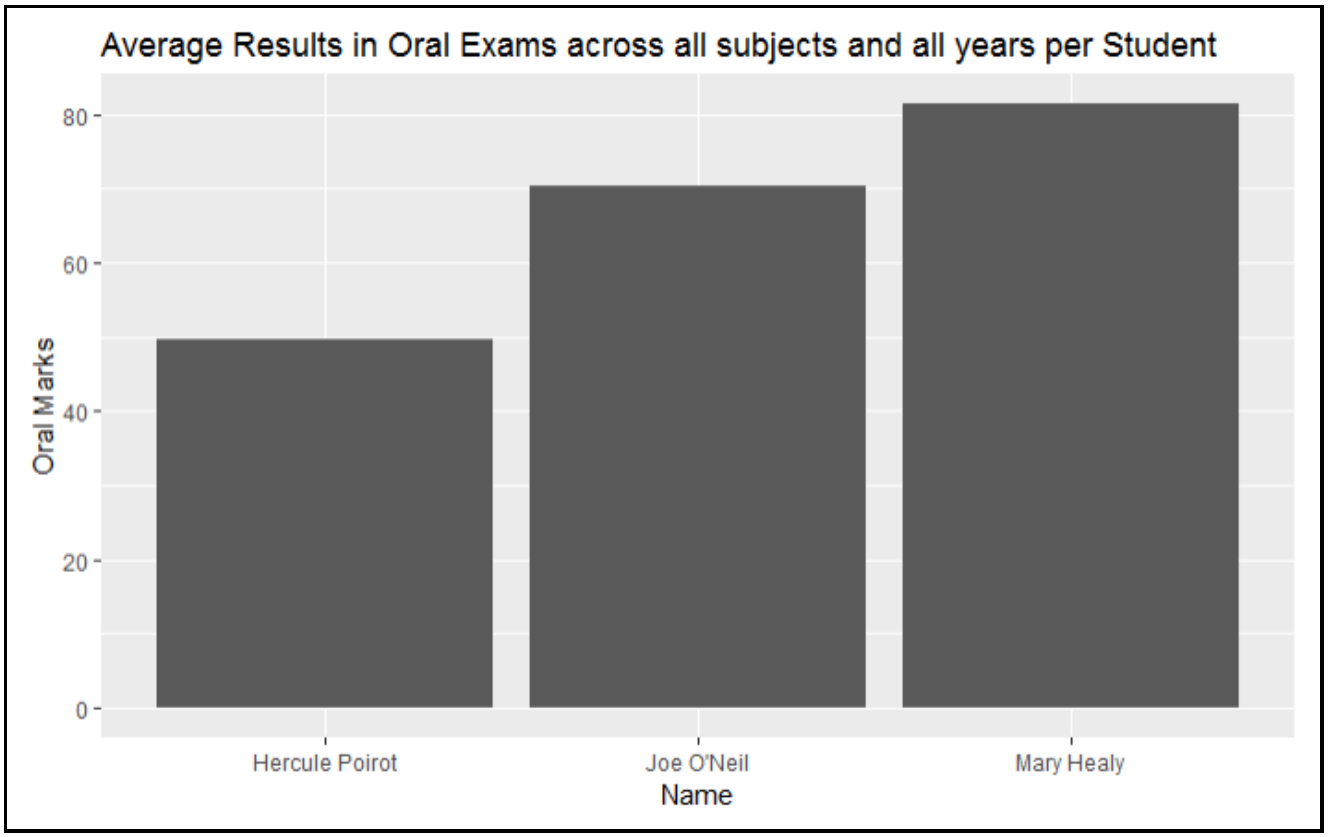
Average results in written exams across all subjects and all years per student

```
namesAvgmarkplot<-ggplot(data=markswritten,aes(x=Name,y=Written_marks))

+labs(y='Written Marks',title='Average Results in Written Exams across all subjects and all years per Student')

namesAvgmarkplot+geom_bar(width = .25,aes(fill=Name),stat = 'identity')+scale_fill_manual(values =
c("Hercule Poirot"='steelblue',"Joe O'Neil"='firebrick',"Mary Healy"='darkgreen'))
```



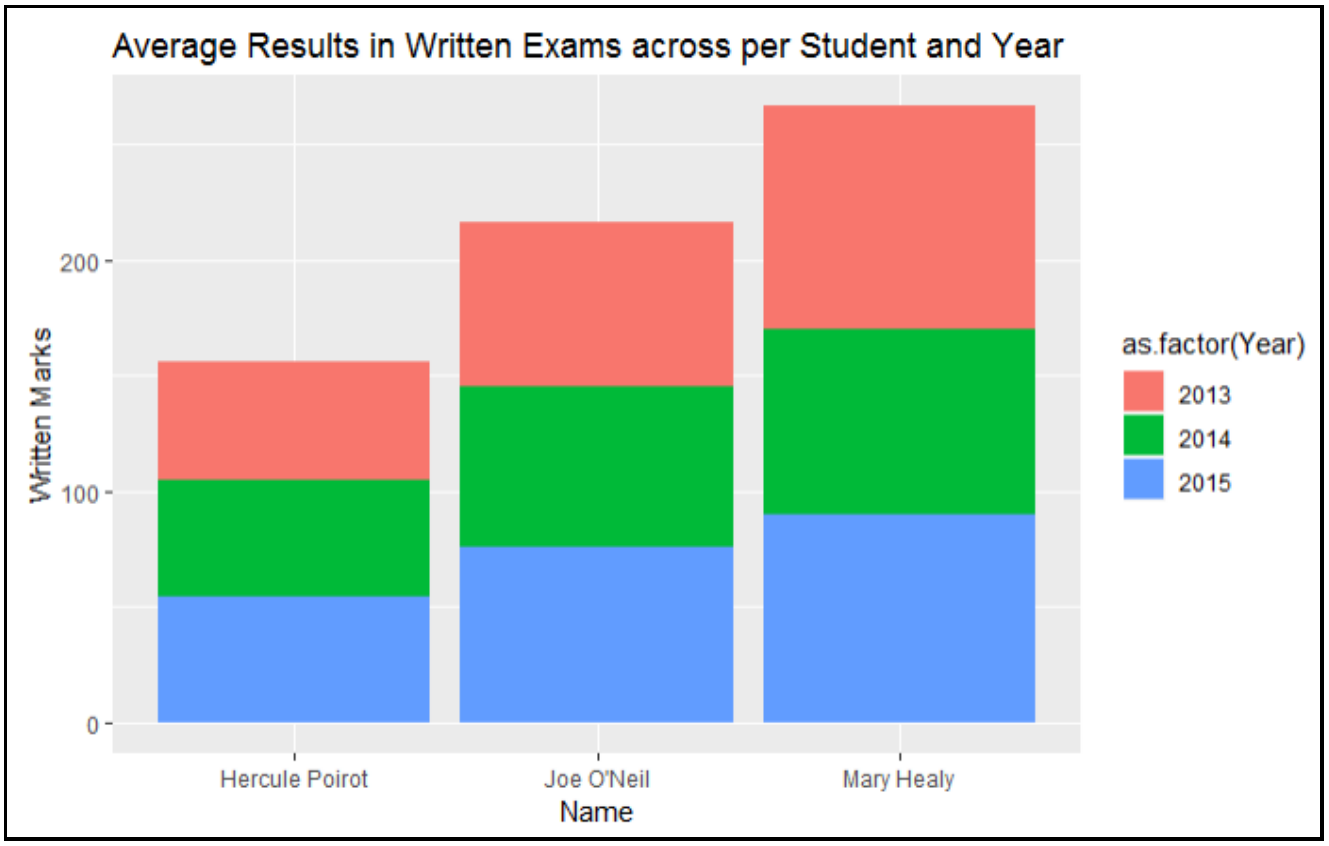
Average results in oral exams across all subjects and all years per student

```
marksoral<-aggregate(data=studentresult,Mark_Oral~Name,FUN = mean)

marksoral

namesAvgoralmarkplot<-ggplot(data=marksoral,aes(x=Name,y=Mark_Oral)) +labs(y='Oral
Marks',title='Average Results in Oral Exams across all subjects and all years per Student')

namesAvgoralmarkplot+geom_bar(stat = 'identity')
```



Average results in the written exams per student and year

```
marks_student<-sqldf("select Year,Name,avg(Mark_Written) as Written_Marks from studentresult group by Year,Name")

marks_student

ggplot(data = marks_student,aes(x=Name,y=Written_Marks,fill=as.factor(Year)))+

  labs(y='Written Marks',title='Average Results in Written Exams across per Student and Year')+

  geom_bar(stat = 'identity')
```




```
# Creating custom functions
getAge<-function(d){
  now<-as.Date(Sys.Date(),format='%d-%m-%Y')
  then<-as.Date(d,format='%d-%m-%Y')
  result<-now-then
  return(round(as.numeric(result/365)))
}
studentresult$age<-getAge(studentresult$DOB)
View(studentresult)
```

Total marks (oral plus written divided by two) for each student for each subject

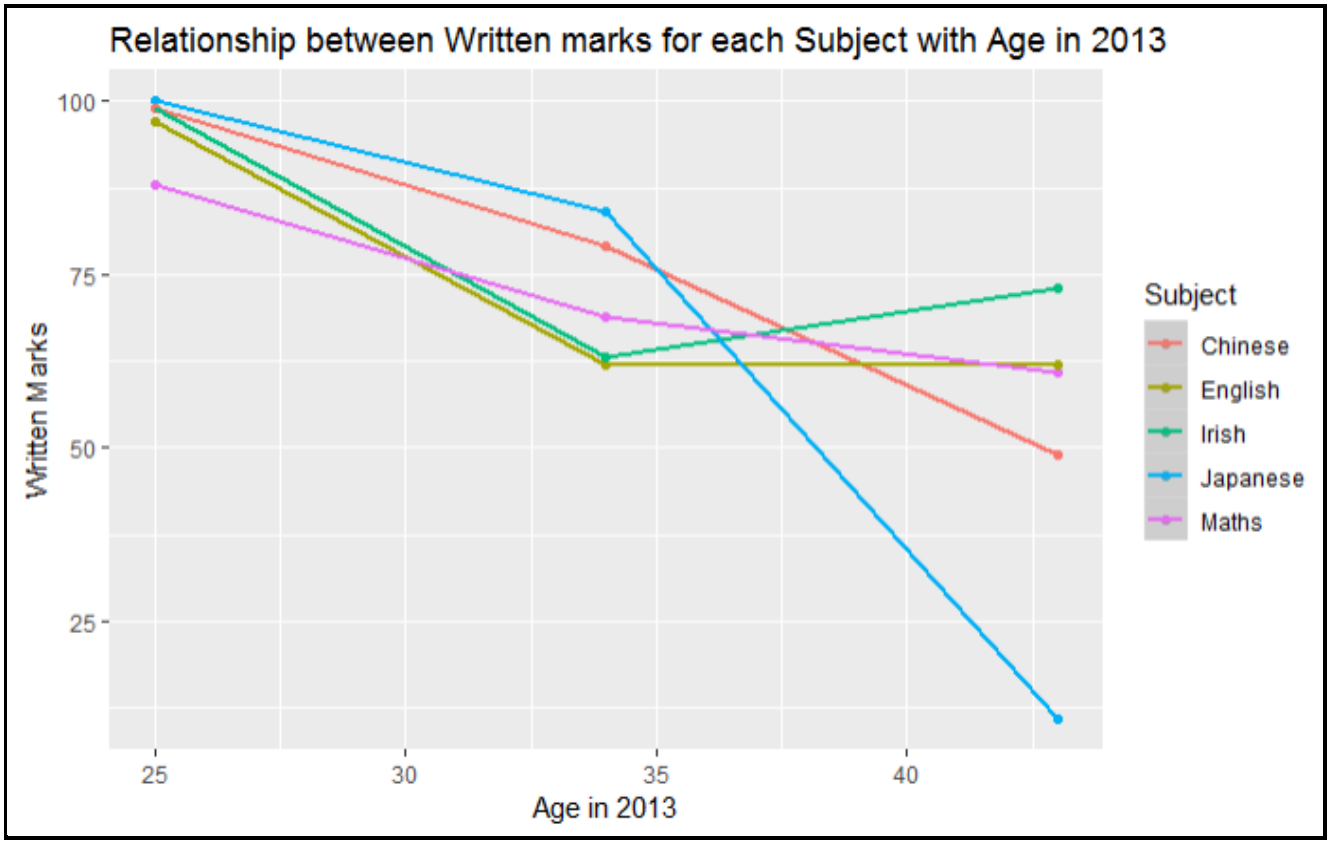
```
total_marks<-sqldf("select Name,Subject,Year,(Mark_Written+Mark_Oral)/2 as Total_Marks from studentresult
group by Name,Subject,Year")
total_marks
ggplot(data=total_marks,aes(x=Name,y=Total_Marks))+
    labs(y='Total Marks', title='Total marks for each Student for each Subject')+
    geom_bar(stat = 'identity',aes(fill=Name))+
```

```
  facet_grid(Year~Subject)
```

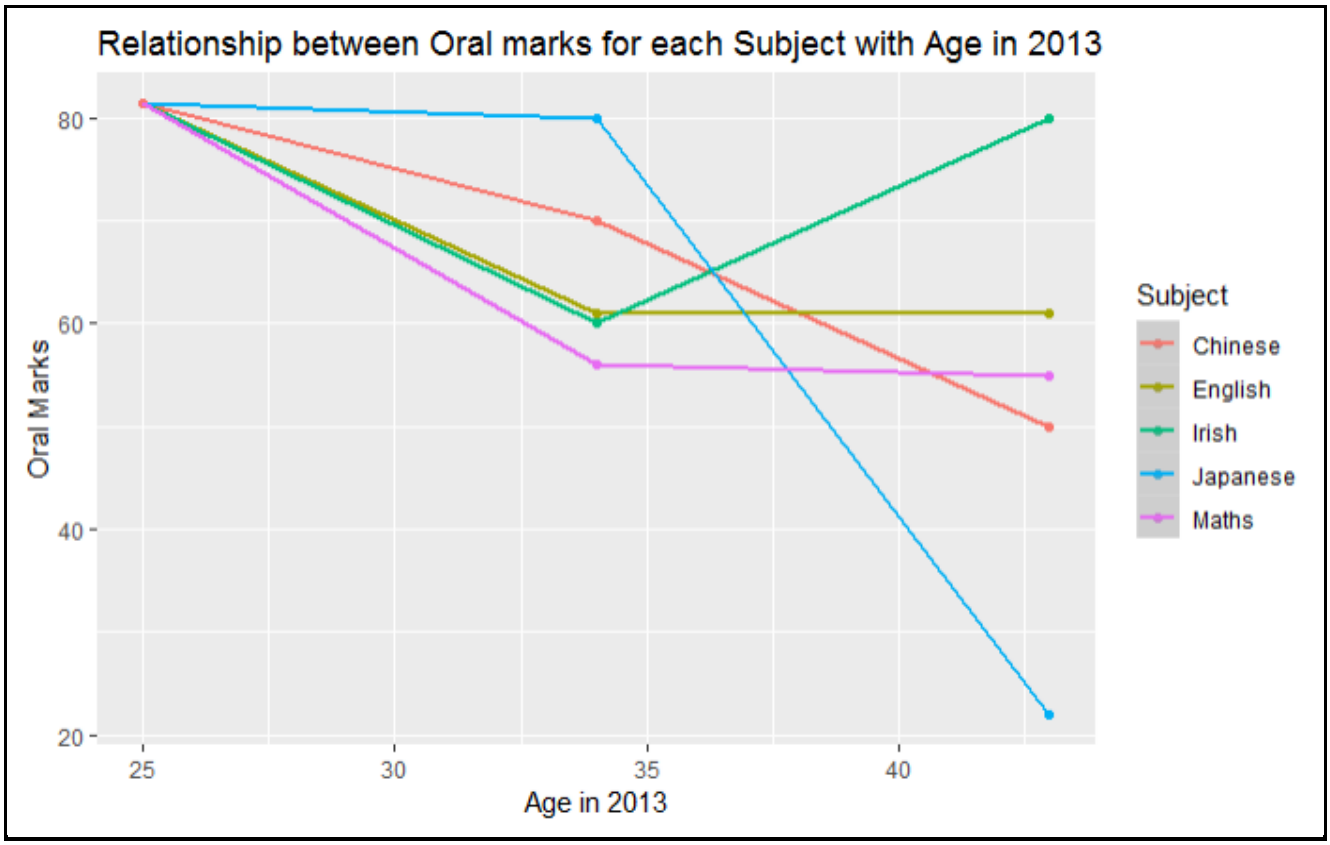
Exploring Relationship between age and mark

```
getAge_new<-function(d){
  ref<-as.Date('01-01-2013',format='%d-%m-%Y')
  then<-as.Date(d,format='%d-%m-%Y')
  result<-ref-then
  return(round(as.numeric(result/365)))
}


# create subsets of data for different years
relation_2013<-subset.data.frame(studentresult,studentresult$Year=='2013')
relation_2013$age2013<-getAge_new(relation_2013$DOB)
View(relation_2013)


relation_2014<-subset.data.frame(studentresult,studentresult$Year=='2014')
relation_2014$age2014<-getAge_new(relation_2014$DOB)+1
View(relation_2014)
relation_2015<-subset.data.frame(studentresult,studentresult$Year=='2015')
relation_2015$age2015<-getAge_new(relation_2015$DOB)+2
View(relation_2015)
```

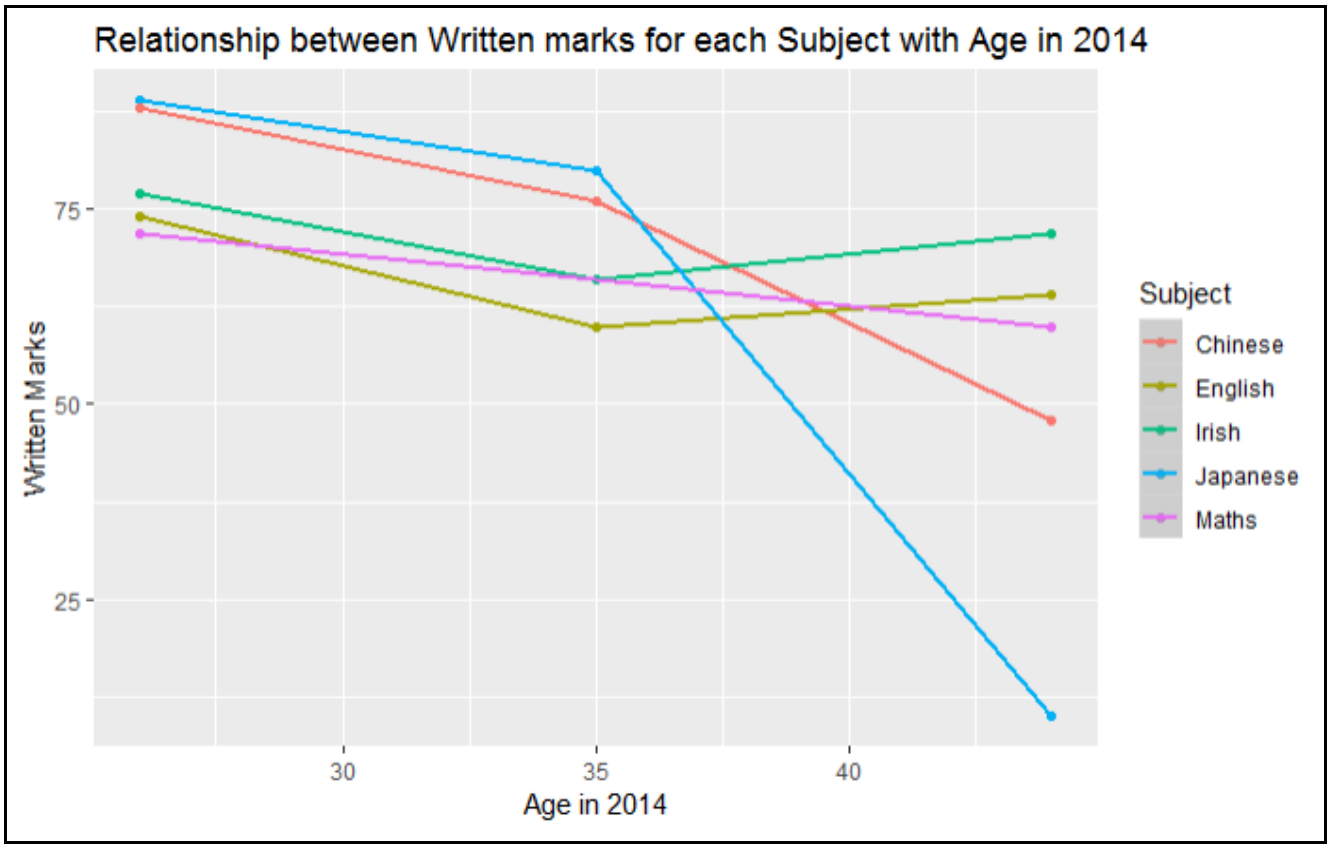
```
# Relation of Written Marks with Age in 2013
ggplot(data = relation_2013)+
    labs(x='Age in 2013',y='Written Marks', title='Relationship between Written marks for each Subject with Age in 2013')+
    geom_point(stat='identity',aes(x=age2013,y=Mark_Written,colour=Subject))+
    geom_smooth(stat='identity',aes(x=age2013,y=Mark_Written,colour=Subject))
```



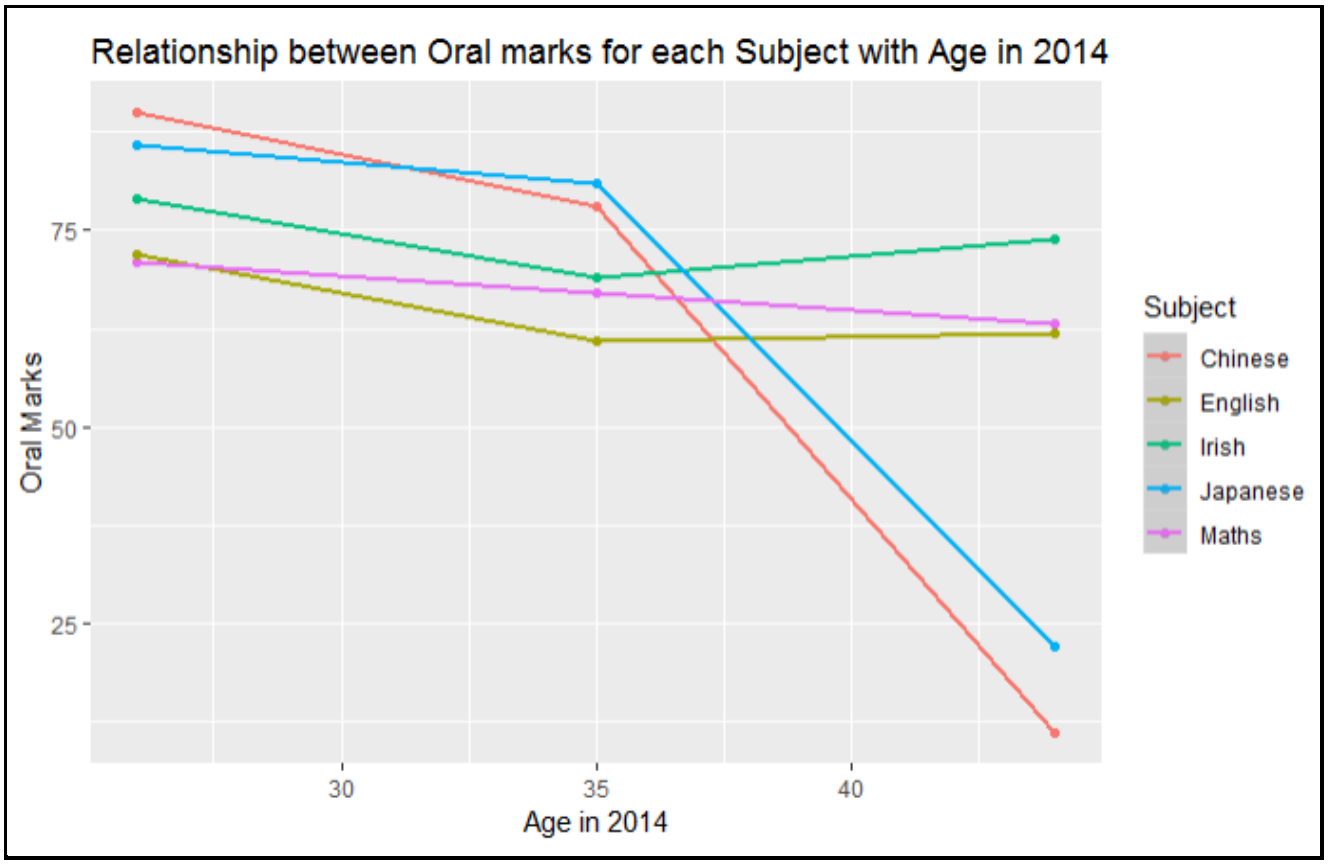

```
# Relation of Oral Marks with Age in 2013
ggplot(data = relation_2013)+
  labs(x='Age in 2013',y='Oral Marks', title='Relationship between Oral marks for each Subject with Age in 2013')+
  geom_point(stat='identity',aes(x=age2013,y=Mark_Oral,colour=Subject))+
  geom_smooth(stat='identity',aes(x=age2013,y=Mark_Oral,colour=Subject))
```



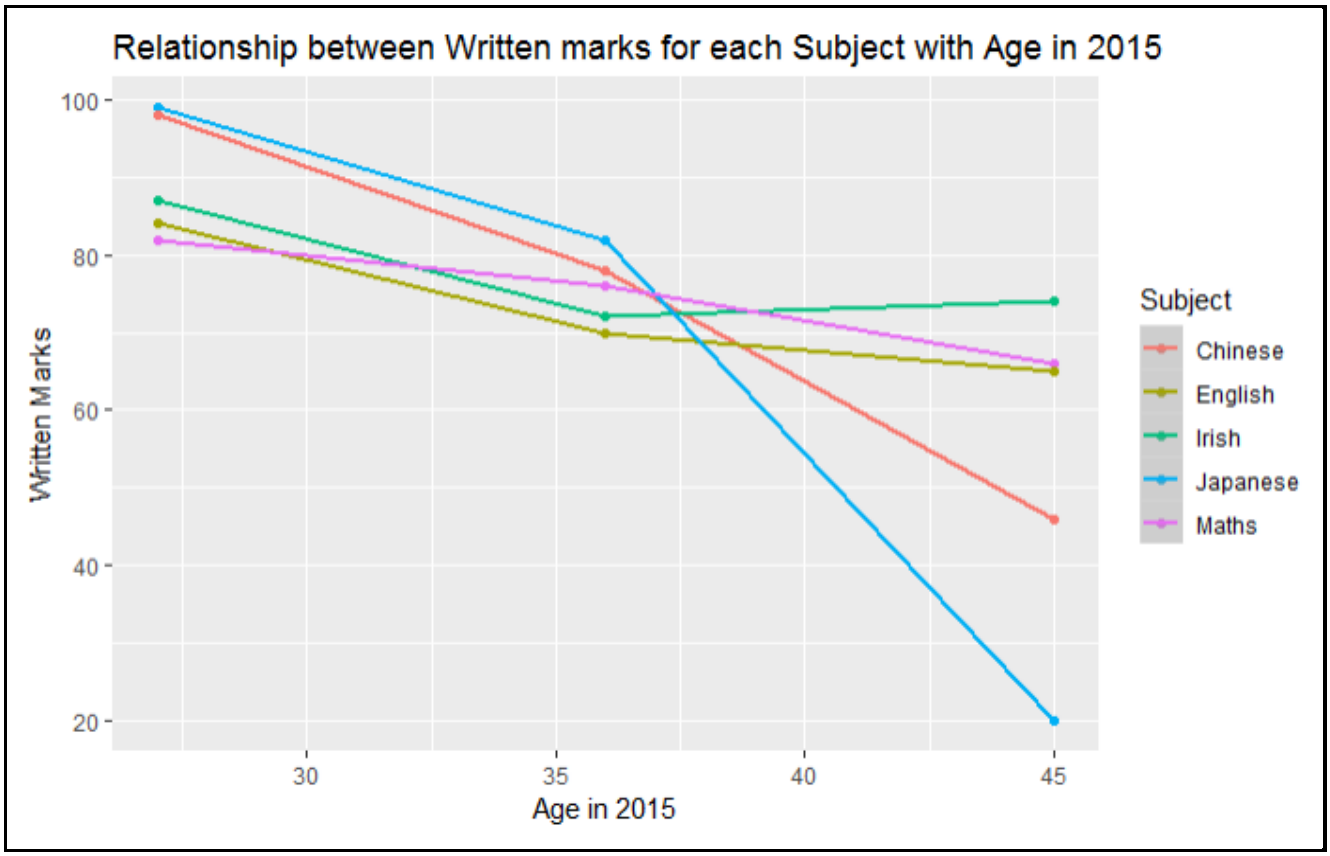

```
# Relation of Written Marks with Age in 2014

ggplot(data = relation_2014)+

  labs(x='Age in 2014',y='Written Marks', title='Relationship between Written marks for each Subject with Age in 2014')+

  geom_point(stat='identity',aes(x=age2014,y=Mark_Written,colour=Subject))+

  geom_smooth(stat='identity',aes(x=age2014,y=Mark_Written,colour=Subject))
```



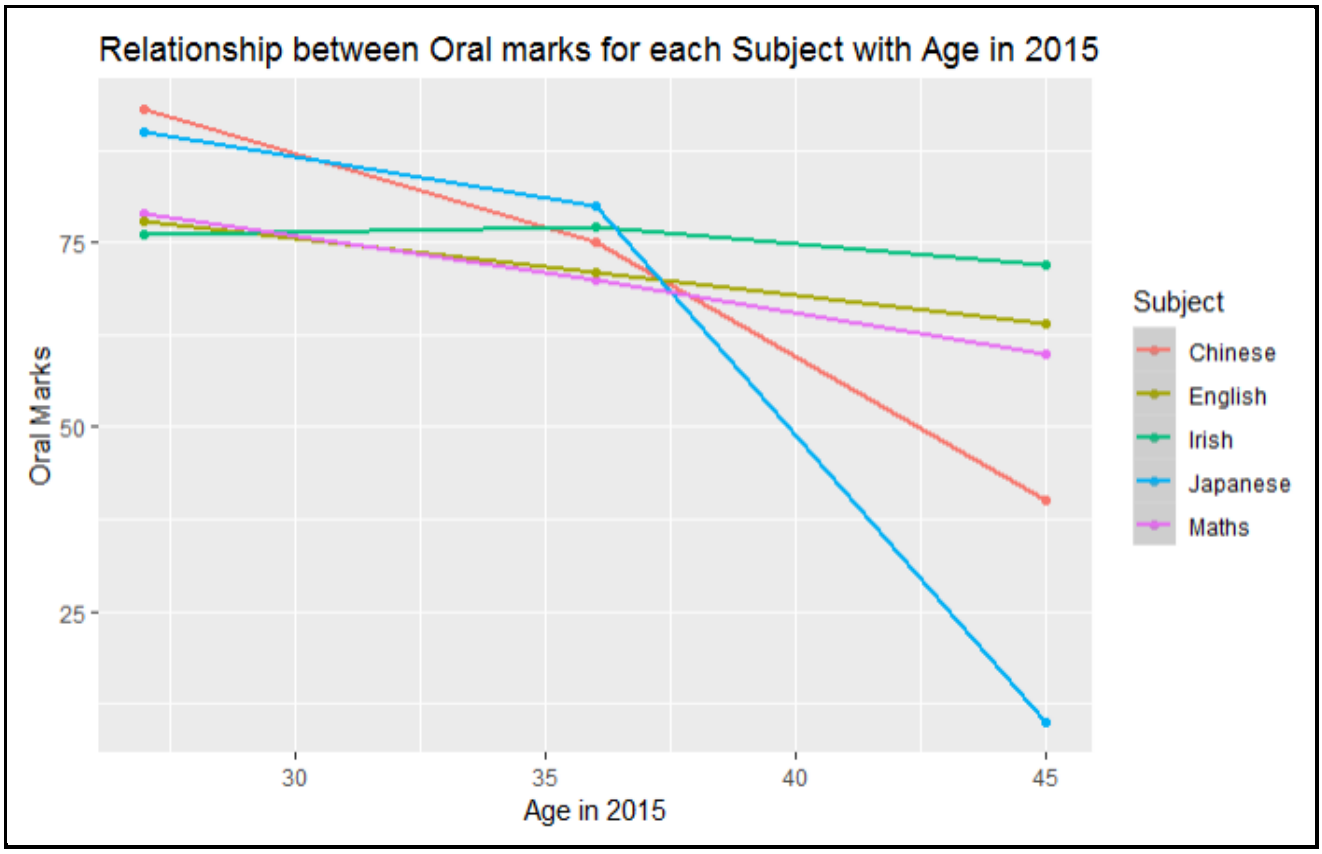

```
# Relation of Oral Marks with Age in 2014
ggplot(data = relation_2014)+
  labs(x='Age in 2014',y='Oral Marks', title='Relationship between Oral marks for each Subject with Age in 2014')+
  geom_point(stat='identity',aes(x=age2014,y=Mark_Oral,colour=Subject))+
  geom_smooth(stat='identity',aes(x=age2014,y=Mark_Oral,colour=Subject))
```




```
# Relation of Written Marks with Age in 2015

ggplot(data = relation_2015)+

  labs(x='Age in 2015',y='Written Marks', title='Relationship between Written marks for each Subject with Age in 2015')+

  geom_point(stat='identity',aes(x=age2015,y=Mark_Written,colour=Subject))+

  geom_smooth(stat='identity',aes(x=age2015,y=Mark_Written,colour=Subject))
```




```
# Relation of Oral Marks with Age in 2015
ggplot(data = relation_2015)+
  labs(x='Age in 2015',y='Oral Marks', title='Relationship between Oral marks for each Subject with Age in 2015')+
  geom_point(stat='identity',aes(x=age2015,y=Mark_Oral,colour=Subject))+
  geom_smooth(stat='identity',aes(x=age2015,y=Mark_Oral,colour=Subject))
```



Did any students do better on their written compared with their oral (or vice versa)

markscomp<-sqldf('select Name,sum(Mark_Written) as Total_Written_marks,sum(Mark_Oral) as Total_Oral_marks from studentresult group by Name')

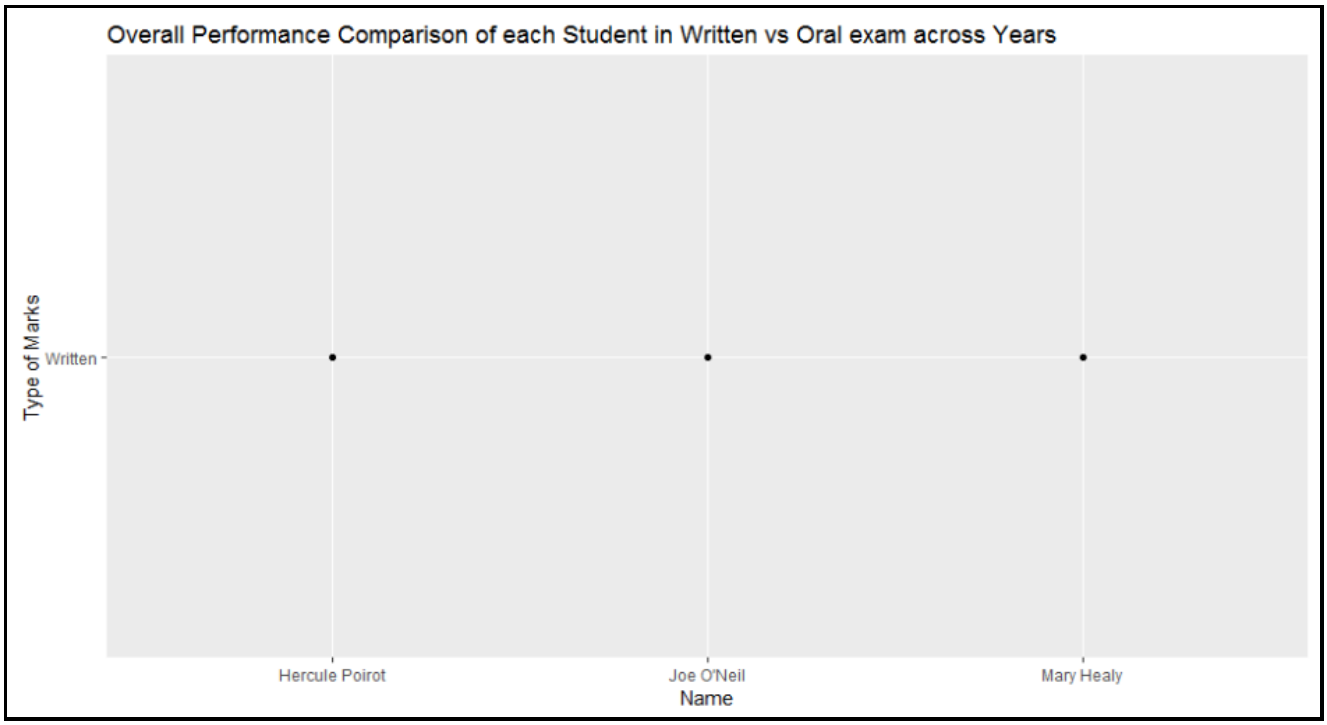
markscomp$compared<-
ifelse(markscomp$Total_Written_marks>markscomp$Total_Oral_marks,'Written','Oral')

markscomp

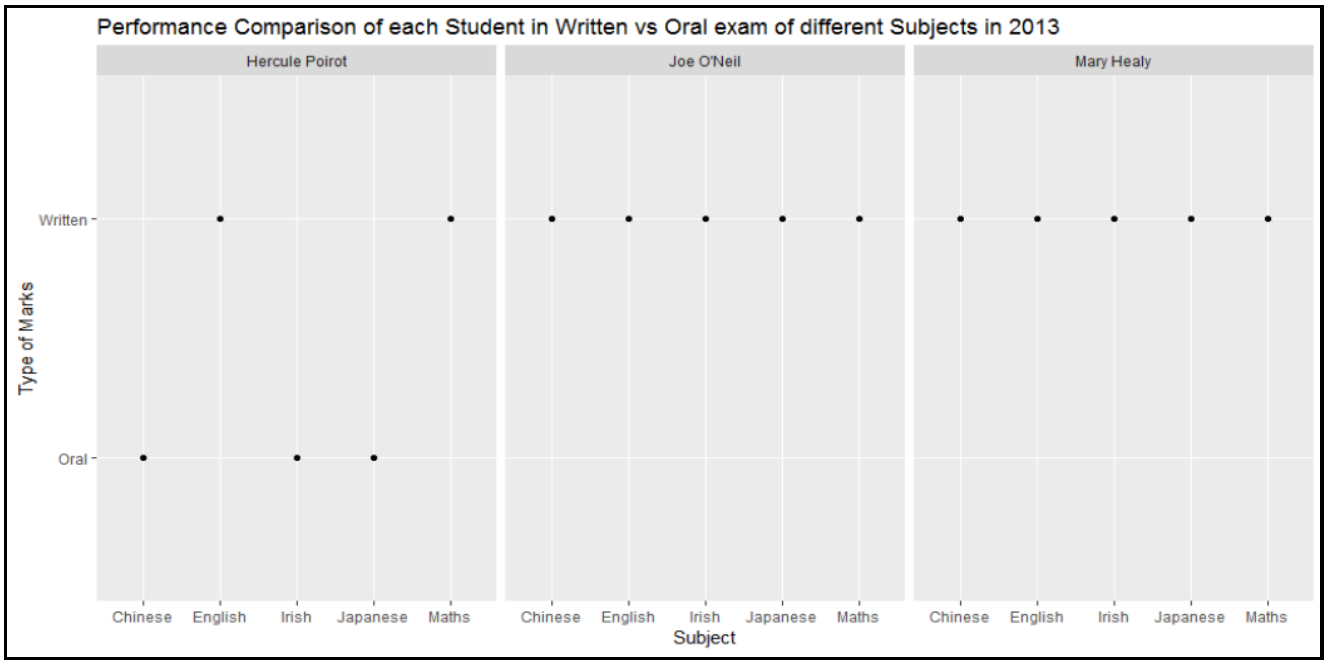
ggplot(data = markscomp)+

labs(x='Name',y='Type of Marks', title='Overall Performance Comparison of each Student in Written vs Oral exam across Years')+

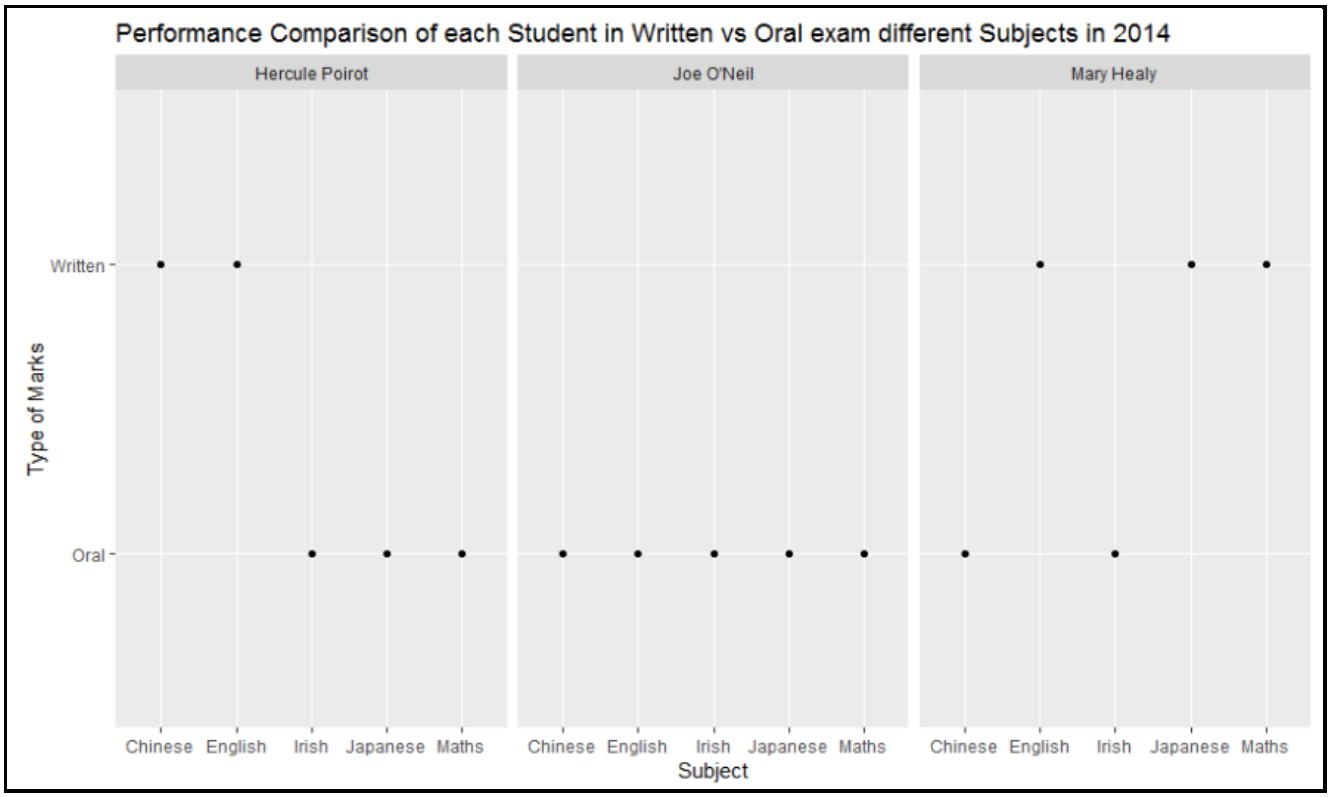
geom_point(stat='identity',aes(x=Name,y=compared),width=2)



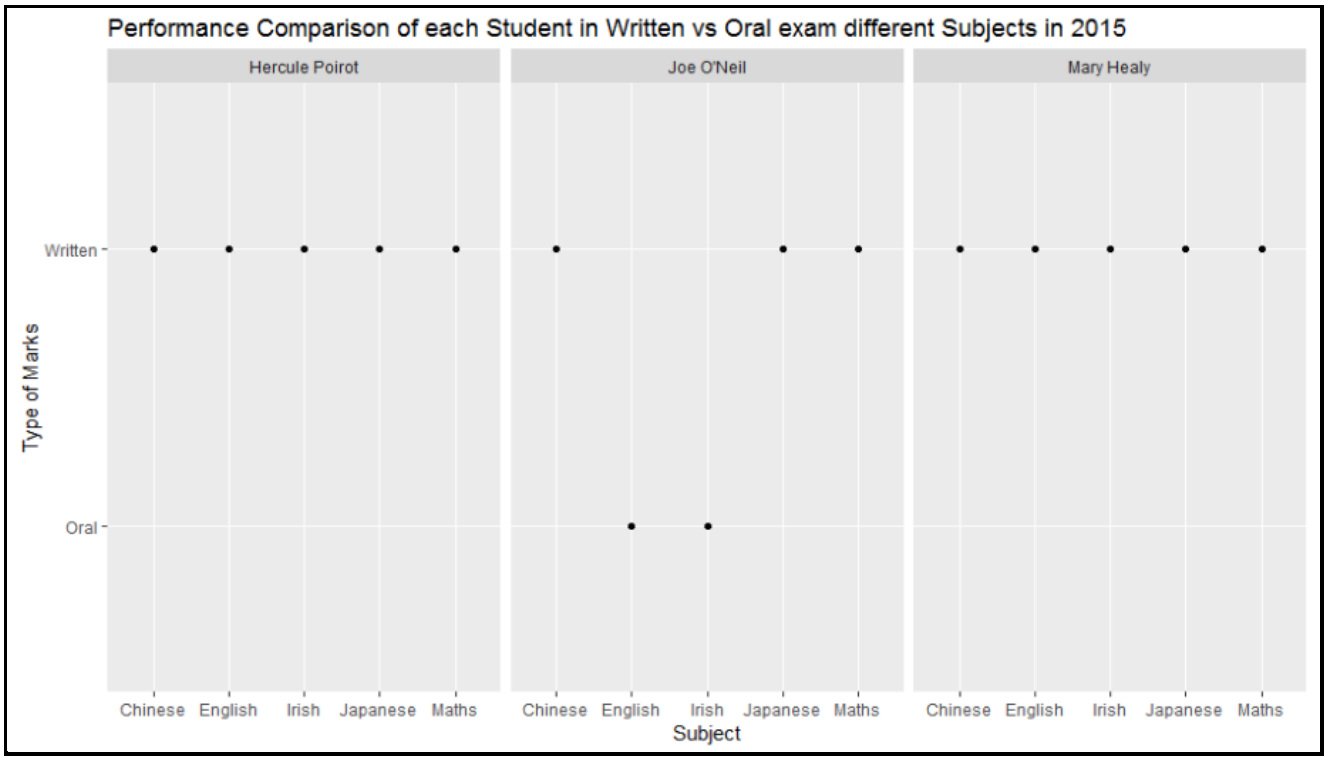

```
# comparison of written and oral marks based on Subject for year 2013

relation_2013$compared<-ifelse(relation_2013$Mark_Written>relation_2013$Mark_Oral,'Written','Oral')

View(relation_2013)

ggplot(data = relation_2013)+

  labs(x='Subject',y='Type of Marks', title='Performance Comparison of each Student in Written vs Oral exam of different Subjects in 2013')+

  geom_point(stat='identity',aes(x=Subject,y=compared),width=2)+

  facet_wrap(.~Name)
```




```
# comparison of written and oral marks based on Subject for year 2014

relation_2014$compared<-ifelse(relation_2014$Mark_Written>relation_2014$Mark_Oral,'Written','Oral')

View(relation_2014)

ggplot(data = relation_2014)+

  labs(x='Subject',y='Type of Marks', title='Performance Comparison of each Student in Written vs Oral exam different Subjects in 2014')+

  geom_point(stat='identity',aes(x=Subject,y=compared),width=2)+

  facet_wrap(.~Name)
```




```
# comparison of written and oral marks based on Subject for year 2015

relation_2015$compared<-ifelse(relation_2015$Mark_Written>relation_2015$Mark_Oral,'Written','Oral')

View(relation_2015)

ggplot(data = relation_2015)+

  labs(x='Subject',y='Type of Marks', title='Performance Comparison of each Student in Written vs Oral exam different Subjects in 2015')+

  geom_point(stat='identity',aes(x=Subject,y=compared),width=2)+

  facet_wrap(.~Name)
```



Subject that obtained the best results on average

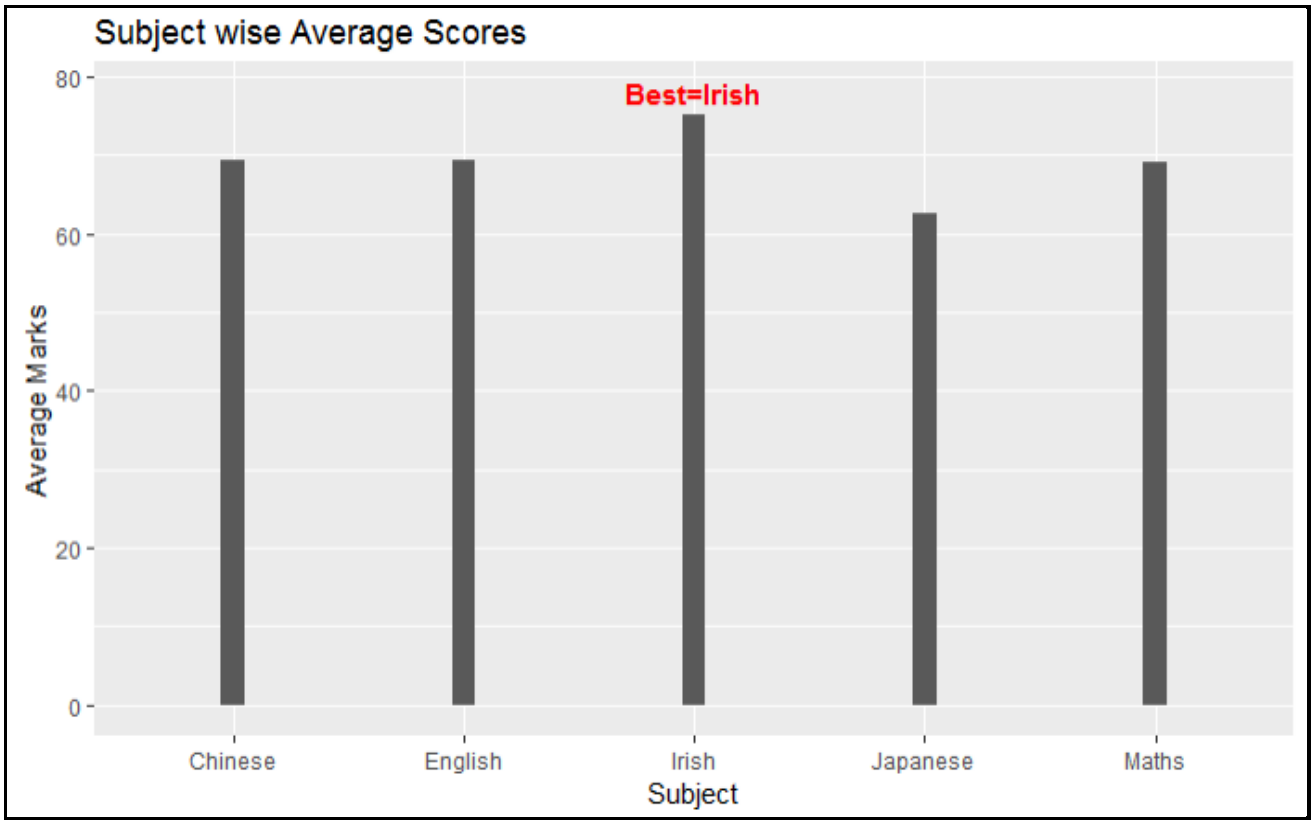
avg_marks<-sqldf('select Subject,avg(Mark_Written+Mark_Oral)/2 as average_marks from studentresult group by Subject')

avg_marks

ggplot(data=avg_marks,aes(x=Subject,y=average_marks))+

labs(y='Average Marks', title='Subject wise Average Scores')+

geom_bar(stat = 'identity',width = .1) +annotate("text", colour='Red',label='Best=Irish', x = 3, y = 78, fontface="bold")


Subject wise Average Scores
Best=Irish
80
60
40
20
0
Average Marks
Chinese
English
Irish
Japanese
Maths
Subject