Compare

# DoDownload the data

setwd('C:/Users/VerkhovodTS/Desktop/R')  
f1 <- read.csv2('AGE\_test.csv', header = TRUE, encoding = 'UNICOD')  
f2 <- read.csv2('AGE\_fit.csv', header = TRUE, encoding = 'UNICOD')  
f2 <- f2[-1]  
f <- dplyr::bind\_cols(f1, f2)  
f <- f[,-1]  
head(f)

## GDP Unemployment Average.life.expectancy Constitutional.form  
## 1 26.75875 1.609438 82 2  
## 2 24.72562 1.791759 74 2  
## 3 28.13180 1.791759 82 1  
## 4 30.14062 1.386294 76 2  
## 5 23.84292 2.397895 81 2  
## 6 26.51932 1.791759 81 1  
## GDP.per.capita Ec.active.population Birth.rate p\_sr p\_mr p\_pr  
## 1 10.769516 51.90916 10 82.47448 82.07160 81.86583  
## 2 8.659040 53.66262 11 73.12811 75.76733 73.77011  
## 3 10.717723 55.18000 10 82.24511 81.92563 81.74037  
## 4 9.091444 56.61009 12 75.04304 76.62995 75.85982  
## 5 10.178806 51.80074 11 79.85848 80.05066 80.18878  
## 6 10.953277 51.80438 11 83.28827 82.23344 82.27557  
## p\_dt p\_rf  
## 1 81.87500 82.16667  
## 2 72.13636 73.45000  
## 3 81.87500 81.79167  
## 4 76.79310 77.40000  
## 5 81.87500 81.50000  
## 6 81.87500 82.56667

# Compare models

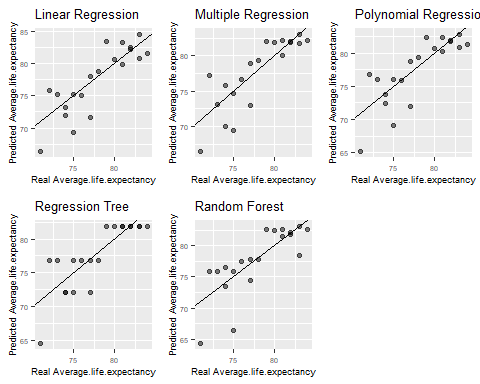
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.6.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

g\_sr <- ggplot(f, aes(x=Average.life.expectancy, y=p\_sr)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Linear Regression", x="Real Average.life.expectancy", y="Predicted Average.life.expectancy") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_mr <- ggplot(f, aes(x=Average.life.expectancy, y=p\_mr)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Multiple Regression", x="Real Average.life.expectancy", y="Predicted Average.life.expectancy") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_pr <- ggplot(f, aes(x=Average.life.expectancy, y=p\_pr)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Polynomial Regression", x="Real Average.life.expectancy", y="Predicted Average.life.expectancy") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")   
  
g\_dt <- ggplot(f, aes(x=Average.life.expectancy, y=p\_dt)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Regression Tree", x="Real Average.life.expectancy", y="Predicted Average.life.expectancy") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
g\_rf <- ggplot(f, aes(x=Average.life.expectancy, y=p\_rf)) +   
 geom\_abline(intercept=0, slope=1) +  
 geom\_point(alpha=0.5) + labs(title="Random Forest", x="Real Average.life.expectancy", y="Predicted Average.life.expectancy") +   
 theme(plot.title=element\_text(size=10), axis.title.x=element\_text(size=7), axis.title.y=element\_text(size=7), axis.text.x=element\_text(size=5), axis.text.y=element\_text(size=5)) + theme(legend.position="none")  
  
gridExtra::grid.arrange(g\_sr,g\_mr,g\_pr,g\_dt,g\_rf,ncol=3)

 # Calc prediction error and visualize it

sr <- mean ((f$Average.life.expectancy - f$p\_sr) ^ 2)   
mr <- mean ((f$Average.life.expectancy - f$p\_mr) ^ 2)   
pr <- mean ((f$Average.life.expectancy - f$p\_pr) ^ 2)   
dt <- mean ((f$Average.life.expectancy - f$p\_dt) ^ 2)   
rf <- mean ((f$Average.life.expectancy - f$p\_rf) ^ 2)   
mse <- data.frame(sr,mr,pr,dt,rf)   
head(mse)

## sr mr pr dt rf  
## 1 7.557154 6.948426 7.944063 7.20737 9.949549

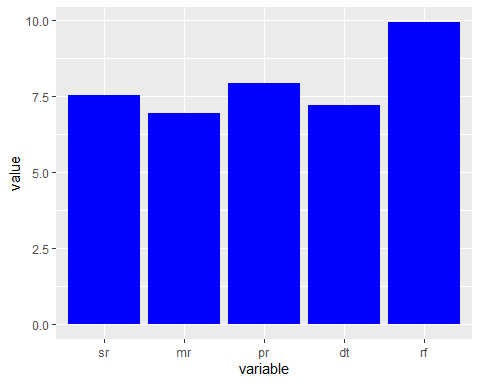
mse1 <- reshape::melt.data.frame(mse)

## Using as id variables

head(mse1)

## variable value  
## 1 sr 7.557154  
## 2 mr 6.948426  
## 3 pr 7.944063  
## 4 dt 7.207370  
## 5 rf 9.949549

b1 <- ggplot(mse1, aes(x=variable, y=value)) + geom\_bar(stat="summary", fun=mean, fill = 'blue')  
b1

 # Save results

ggsave("plot.jpg", plot=b1 + theme\_classic(), width=20, height=15, units="cm", dpi=600)