HOMEWORK 4

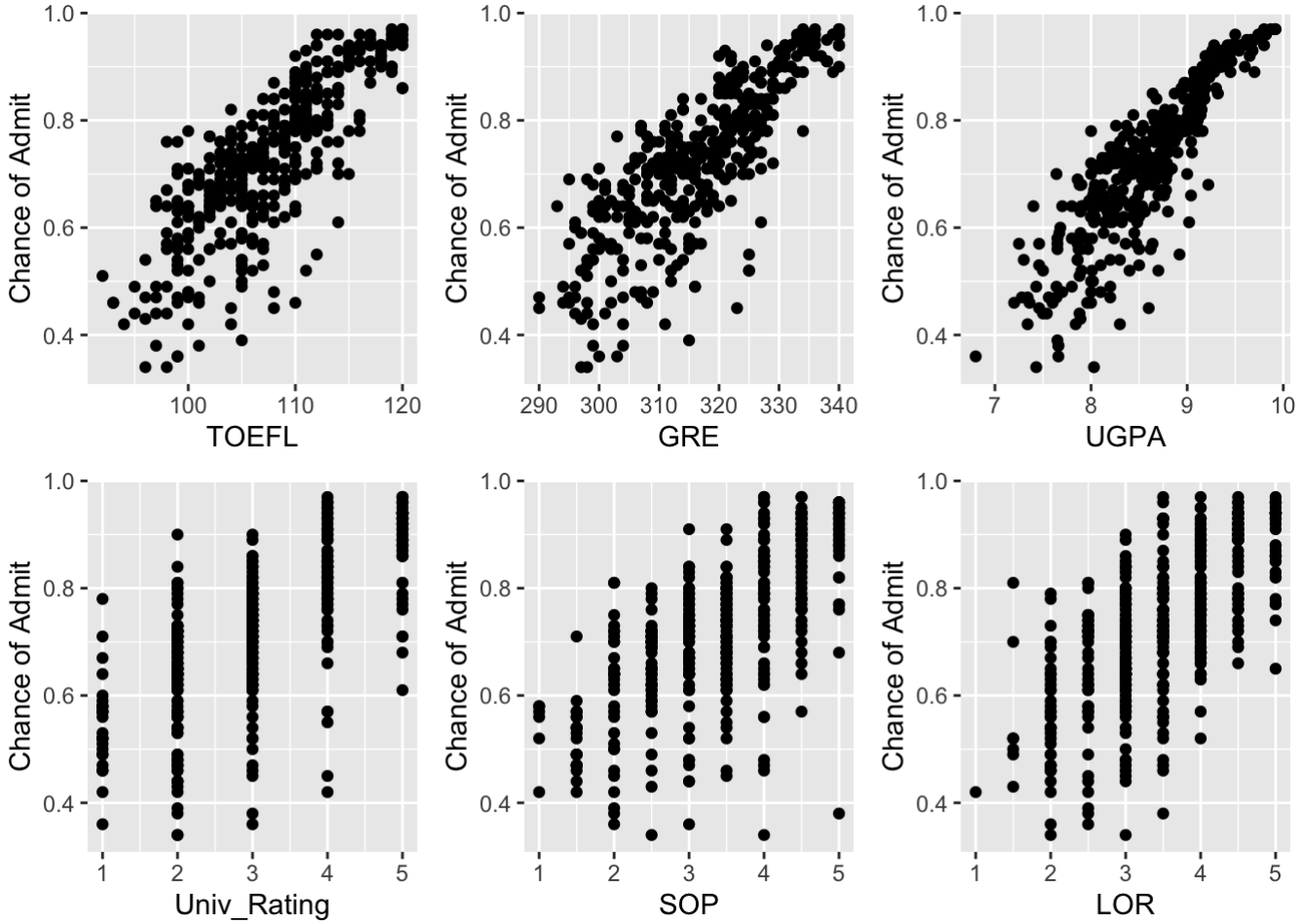
Raul Rodriguez

2023-07-27

```
library(gridExtra)
library(ggplot2)

admissions_data <- read.csv('AdmissionPredict_202305.csv')

col_names <- c('TOEFL', 'GRE' ,'UGPA' ,'Univ_Rating' , 'SOP' , 'LOR')

create_plot <- function(name) {
  ggplot(admissions_data, aes_string(x = name, y = "Chance_of_Admit")) +
    geom_point() +
    labs(x = name, y = "Chance of Admit")
}

#a)
plots <- lapply(col_names, create_plot)
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## ℹ Please use tidy evaluation idioms with `aes()`.
## ℹ See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
do.call(grid.arrange, c(plots, ncol = 3))
```




```
#b)
# I am able to identify a relationship  in TOEFL, GRE, UGPA,
# CHANCE_OF_ADMIT, LOR, SOP, AND Univ_Rating. The type of relationship
# that i see is linear with an upward trend.

#C)
model <- lm(Chance_of_Admit ~ TOEFL + GRE + UGPA + Univ_Rating + SOP + LOR, data = ad
missions_data)
model
```

```
##
## Call:
## lm(formula = Chance_of_Admit ~ TOEFL + GRE + UGPA + Univ_Rating +
##     SOP + LOR, data = admissions_data)
##
## Coefficients:
## (Intercept)        TOEFL          GRE         UGPA  Univ_Rating          SOP
##   -1.413859     0.002753     0.002276     0.119875     0.006062    -0.001961
##         LOR
##    0.022749
```

```
#d)
# All terms are significant (p < 0.05) with the exception of UGPA. This means
# That all values with the exception of UGPA contribute to the prediction of
# Chance_of_Admit.

#e)
new_model <- lm(Chance_of_Admit ~ TOEFL + GRE + Univ_Rating + SOP + LOR, data = admis
sions_data)
new_model
```

```
##
## Call:
## lm(formula = Chance_of_Admit ~ TOEFL + GRE + Univ_Rating + SOP +
##     LOR, data = admissions_data)
##
## Coefficients:
## (Intercept)        TOEFL          GRE  Univ_Rating          SOP          LOR
##   -1.544756     0.006002     0.004553     0.014649     0.005910     0.033745
```

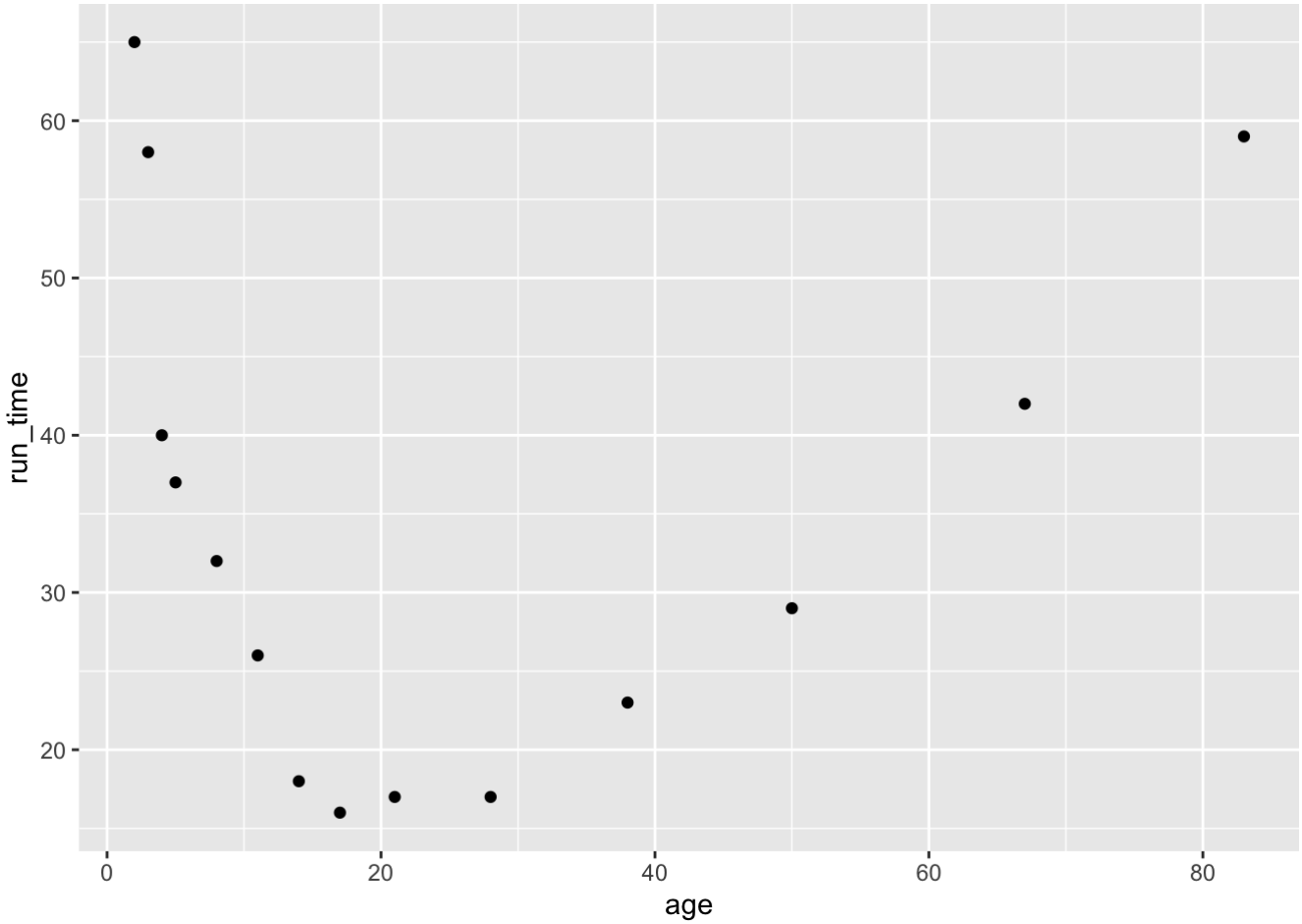
```
#f)

range_values <- sapply(admissions_data[, c("TOEFL", "GRE", "Univ_Rating", "SOP", "LO
R")], range)
range_values
```

```
##      TOEFL GRE Univ_Rating SOP LOR
## [1,]    92 290           1   1   1
## [2,]   120 340           5   5   5
```

```
#g)

TOEFL_score <- 108
GRE_score <- 330
UGPA_score <- 7.8
Univ_Rating <- 4
SOP_score <- 3
LOR_score <- 4

new_student <- data.frame(TOEFL = TOEFL_score, GRE = GRE_score, UGPA = UGPA_score, Un
iv_Rating = Univ_Rating, SOP = SOP_score, LOR = LOR_score)

predicted_chance <- predict(new_model, newdata = new_student)

predicted_chance
```

```
##         1
## 0.8174412
```

```
age = c(2,3,4,5,8,11,14,17,21,28,38,50,67,83)

 run.time = c(65,58,40,37,32,26,18,16,17,17,23,29,42,59)

 data <- data.frame(age = age, run_time = run.time)

 #a)
 # The response variable (dependent) is the run time,
 # as it depends on the age of the runner. Which would
 # make age the explanatory variable (independent)

 #b)
 ggplot(data = data, aes(x = age, y = run_time)) +
   geom_point()
```




```
 # The relationship forms a u-shape. Meaning that the youngest and the oldest have th
e
 # highest running times. While the 'middle' group scores the lowest running times.

 #c)
 model <- lm(run.time ~ age + I(age^2), data = data)

 #d)
 test_person <- data.frame(age = 56)
 predicted_run_time <- predict(model, test_person = new_data)
 predicted_run_time <- predict(model, test_person = new_data)
 predicted_run_time
```

```
##        1        2        3        4        5        6        7        8
## 46.70564 44.97106 43.28589 41.65012 37.03929 32.87313 29.15166 25.87487
##        9       10       11       12       13       14
## 22.19754 17.66447 15.38843 19.17917 36.72874 66.28998
```

```
 #e)

 r_squared <- summary(model)$r.squared
print('percentage: ')
```

```
## [1] "percentage: "
```

```
r_squared * 100
```

```
## [1] 69.96079
```