Data Science - R Programming - Quiz 1 - Coursera

Quiz 1

This is Quiz 1 from Coursera's R Programming class within the Data Science Specialization. This publication is intended as a learning resource, all answers are documented and explained. Datasets are available in R packages.

---

1. The R language is a dialect of which of the following programming languages?

- **S**

Explanation:

R is an open source implementation of S with a revised syntax and an awesome community.

---

2. The definition of free software consists of four freedoms (freedoms 0 through 3). Which of the following is NOT one of the freedoms that are part of the definition? Select all that apply.

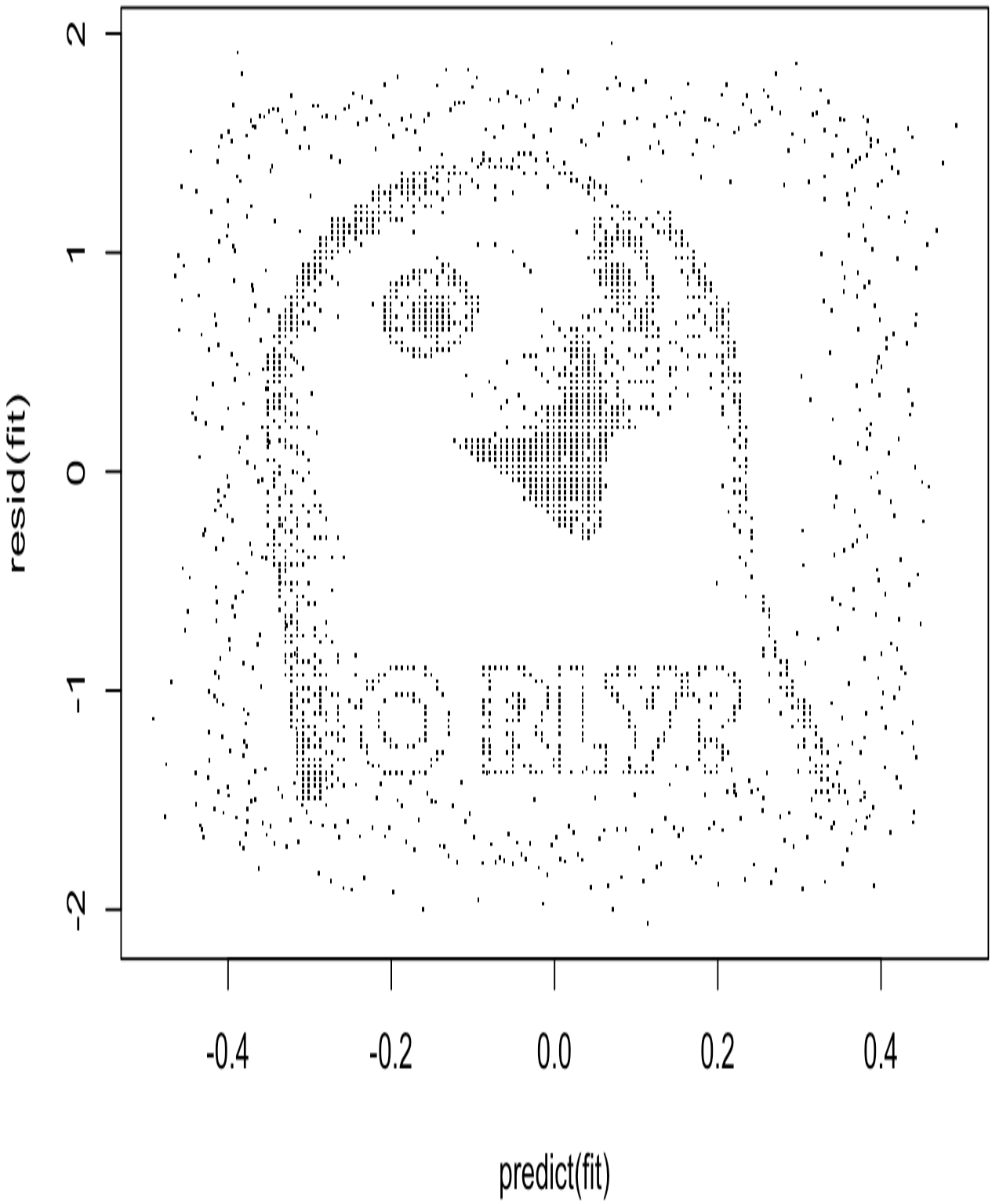
- **The freedom to sell the software for any price.**
- **The freedom to restrict access to the source code for the software.**
- **The freedom to prevent users from using the software for undesirable purposes.**

Explanation:

Yay free software!

```
dat <- read.table('http://www4.stat.ncsu.edu/~stefanski/NSF_Supported/Hidden_Images/orly_owl_files/orly_owl_Lin_4p_5_flat.txt', header = FALSE)

fit <- lm(V1 ~ . - 1, data = dat); plot(predict(fit), resid(fit), pch = '.')
```


resid(fit)
2
1
0
-1
-2
-0.4
-0.2
0.0
0.2
0.4
predict(fit)

3. In R the following are all atomic data types EXCEPT: (Select all that apply)

- **list**
- **array**
- **matrix**
- **data frame**
- **table**

Explanation:

Predicting with the lower and upper bounds of the confidence intervals

```
fit <- lm(mpg~wt,mtcars)
summary(fit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

4. If I execute the expression x <- 4 in R, what is the class of the object `x' as determined by the`class()' function?

- **numeric**

Explanation:

R automatically interprets 4 as a numeric class object.

5. What is the class of the object defined by the expression x <- c(4, “a”, TRUE)?

- **character**

Explanation:

```
x <- c(4, "a", TRUE)
class(x)
## [1] "character"
```

---

6. If I have two vectors x <- c(1,3, 5) and y <- c(3, 2, 10), what is produced by the expression cbind(x, y)?

- **a matrix with 2 columns and 3 rows**

Explanation:

Combine the two vectors as columns.

---

7. A key property of vectors in R is that

- **elements of a vector all must be of the same class**

Explanation:

This is nice for statistical purposes.

---

8. Suppose I have a list defined as x <- list(2, “a”, “b”, TRUE). What does x[[2]] give me? Select all that apply.

- **a character vector containing the letter “a”.**
- **a character vector of length 1.**

Explanation:

Two brackets gives the actual element inside of the list, one bracket gives the list with the element inside.

---

9. Suppose I have a vector x <- 1:4 and a vector y <- 2. What is produced by the expression x + y?

- **a numeric vector with elements 3, 4, 5, 6.**

Explanation:

---

10. Suppose I have a vector x <- c(17, 14, 4, 5, 13, 12, 10) and I want to set all elements of this vector that are greater than 10 to be equal to 4. What R code achieves this? Select all that apply.

- **x[x >= 11] <- 4**
- **x[x > 10] <- 4**

Explanation:

Indexing with a boolean.

---

11. Use the Week 1 Quiz Data Set to answer questions 11-20.

In the dataset provided for this Quiz, what are the column names of the dataset?

- **Ozone, Solar.R, Wind, Temp, Month, Day**

Explanation:

Download, unzip, read, print.

```
dat <- download.file('https://d396qusza40orc.cloudfront.net/rprog/data/quiz1_data.zip', d
estfile ="quizdat.zip")

dat <- unzip("quizdat.zip")

dat <- read.csv("hw1_data.csv")
```

```
names(dat)
```

```
## [1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"
```

---

12. Extract the first 2 rows of the data frame and print them to the console. What does the output look like?

Explanation:

Index

```
dat[1:2,]
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
```

---

13. How many observations (i.e. rows) are in this data frame?

- **153**

Explanation:

Nrow()

```
nrow(dat)
```

```
## [1] 153
```

---

14. Extract the last 2 rows of the data frame and print them to the console. What does the output look like?

Explanation:

Correlation(XY)* SDy/SDx

```
dat[152:153,]
```

```
##     Ozone Solar.R Wind Temp Month Day
## 152    18     131  8.0   76     9  29
```

```
## 153    20     223 11.5   68     9  30
```

---

15. What is the value of Ozone in the 47th row?

- **21**

Explanation:

$ notation is useful

```
dat$Ozone[47]
```

```
## [1] 21
```

---

16. How many missing values are in the Ozone column of this data frame?

- **37**

Explanation:

Is NA return T/F values which can be summed to get a count of NAs.

```
sum(is.na(dat$Ozone))
```

```
## [1] 37
```

---

17. What is the mean of the Ozone column in this dataset? Exclude missing values (coded as NA) from this calculation.

- **42.1**

Explanation:

na.rm is a great option for calculation where NAs might interfere

```
mean(dat$Ozone, na.rm=TRUE)
```

```
## [1] 42.12931
```

---

18. Extract the subset of rows of the data frame where Ozone values are above 31 and Temp values are above 90. What is the mean of Solar.R in this subset?

- **212.8**

Explanation:

Which give index of booleans, $ selects columns.

```
mean(dat[which(dat$Ozone >31 & dat$Temp > 90),]$Solar.R)
```

```
## [1] 212.8
```

---

19. What is the mean of “Temp” when “Month” is equal to 6?

- **79.1**

Explanation:

Same as above

```
mean(dat[which(dat$Month == 6),]$Temp)
```

```
## [1] 79.1
```

---

20. Let the slope having fit Y as the outcome and X as the predictor be denoted as β1. Let the slope from fitting X as the outcome and Y as the predictor be denoted as γ1. Suppose that you divide β1 by γ1; in other words consider β1/γ1. What is this ratio always equal to?

- **115**

Explanation:

Need to remove NA for this.

```
max(dat[which(dat$Month == 5),]$Ozone, na.rm = TRUE)
```

```
## [1] 115
```
