Statistics with Recitation — Quiz 3

October 14, 2025

Name: _____ **ID:** _____

1. (**2 points**) What is the output of the following R expression?

```
score <- c(1, 2, 3, 4)
my_vec <- c("Hello World!", 1, TRUE, score)
class(my_vec)
```

   A. integer

   B. logical

   C. numeric

   D. character

2. (**2 points**) Which statement about R data.frame objects is FALSE?

   A. A data frame is a list of equal-length vectors.

   B. Columns in a data frame can have different lengths.

   C. Each column can have different data types.

   D. We can use `data.frame()` function to create such objects.

3. (**2 points**) If we want to use a package `ggplot2` that hasn't been installed, which of the following R code solves the problem?

   A. `library("ggplot2")`

   B.
   ```
   install.packages(ggplot2)
   library("ggplot2")
   ```

   C.
   ```
   install.packages("ggplot2")
   library(ggplot2)
   ```

   D.
   ```
   install.packages(ggplot2)
   library(ggplot2)
   ```

4. (**2 points**) If we want to produce a scatter plot in R, which function should we use?

   A. `geom_text()`

   B. `geom_boxplot()`

   C. `geom_point()`

   D. `geom_scatter()`

5. (**2 points**) Which of the following statements is TRUE?

   A. `setwd("C:\\your\\working\\directory")` works both on Windows and Mac.

   B. `/` is an escape character in R.

   C. `getwd()` function shows your home directory.

   D. `setwd(getwd())` works without any errors.

6. (**2 points**) What is the output of the following R expression?

```
id <- 1:4
score <- c(30, 86, 50, 95)
id_score <- matrix(c(id, score), nrow = 2)
print(id_score)
```

A.

```
     [,1] [,2]
[1,]    1   30
[2,]    2   86
[3,]    3   50
[4,]    4   95
```

B.

```
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]   30   86
[4,]   50   95
```

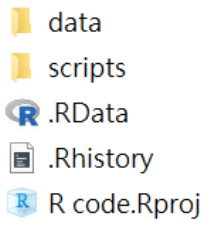
C.

```
      [,1] [,2] [,3] [,4]
[1,]     1    2    3    4
[2,]    30   86   50   95
```

D.

```
      [,1] [,2] [,3] [,4]
[1,]     1    3   30   50
[2,]     2    4   86   95
```

7. (**2 points**) Suppose Alice creates a RStudio project in the TA_Session Folder, and the folder structure is given in the picture. If Alice opens a script `wk_1.R` in the scripts folder, and she wants to import the `county.csv` file from the data folder, which of the following R code will do the job?



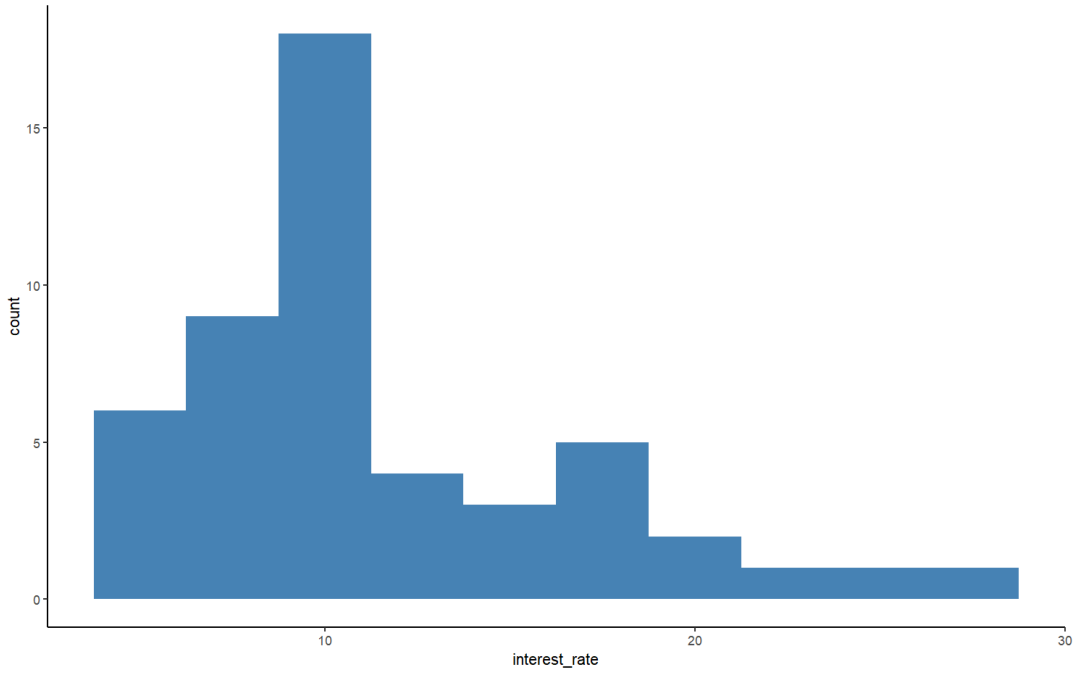
A. `county <- read.csv("county.csv")`

B. **`county <- read.csv("data/county.csv")`**

C. `county <- read.csv("../county.csv")`

D. `county <- read.csv("../data/county.csv")`

8. (**2 points**) Suppose Grimm wants to plot the distribution of value in `interest_rate` of loan data.frame. Which of the following R code would most likely produce the following graph?



A.

```
ggplot(loan, aes(x = interest_rate)) +
  geom_histogram(binwidth = 2.5, fill = "steelblue")
```

B.

```
ggplot(loan, aes(x = interest_rate, y = count)) +
  geom_histogram(binwidth = 2.5, fill = "steelblue")
```

C.

```
ggplot(loan, aes(x = "interest_rate")) +
  geom_histogram(binwidth = 2.5, fill = "steelblue")
```

D.

```
ggplot(loan, aes(x = "interest_rate", y = "count")) +
  geom_histogram(binwidth = 2.5, fill = "steelblue")
```

9. (**2 points**) Which of the following statements are TRUE?

   A. `stackdir` is an argument for the `geom_dotplot()` function that assigns the axis along which the points stack.

   B. We can use `table(..., addmargins = ...)` to add a row of column sums for the table.

   C. The output of `class(ls())` is `"character"`.

   D. The width of boxes in `box_plot()` can not be customized.

10. (**2 points**) What is the output of the following R code?

```
score_summary <- function(x, pass_cut) {
  pass <- 0
  bonus <- 0
  for (i in seq_along(x)) {
    v <- x[i]
    if (v < 50) {

    } else if (v >= 90) {
      bonus <- bonus + 2
    } else {
      bonus <- bonus + 1
    }
    if (v >= pass_cut) pass <- pass + 1
  }
  return(c(pass, bonus))
}

scores <- c(95, 60, 84, 49, 91, 90)
print(score_summary(scores, pass_cut = 60))
```

A.

```
[1] 5 8
```

B.

```
[1] 5 7
```

C.

```
[1] 4 8
```

D.
```
[1] 5 7
```

11. (**2 points**) Which of the following statements about the following list is **TRUE**?

```
info <- list(
  student = "Alice",
  scores = list(mid = 80, final = 90),
  passed = TRUE
)
```

A. `info["scores"]` extracts the numeric vector `c(80, 90)` directly.

B. `info$scores[["mid"]]` and `info[["scores"]]["mid"]` are equivalent.

C. `info[["scores"]][["mid"]]` extracts the value `80`.

D. `info["scores"]["mid"]` correctly returns the value `80`.

12. (**2 points**) Which of the following statements about random number generators is FALSE?

A. The distribution generated from `rbinom(n = 100, size = 30, prob = 0.2)` is most likely to be left-skewed.

B. The default parameters of `rnorm()` set the mean to 0 and the standard deviation to 1.

C. We use `set.seed()` to ensure that the same random values are generated every time the code is run.

D. As the argument `n` becomes larger, the resulting distribution appears smoother.

13. (**2 points**) What is the output of the following R code?

```
x <- c(TRUE, FALSE, TRUE, TRUE, FALSE)
result <- cumsum(as.numeric(x))
print(result)
```

A.
```
[1] 0 1 2 3 3
```

B.
```
[1] 1 1 2 3 3
```

C.
```
[1] 1 0 1 1 0
```

D.
```
[1] TRUE FALSE TRUE TRUE FALSE
```

14. (**2 points**) Which of the following statements about `paste()`, `paste0()`, and `as.Date()` is FALSE?

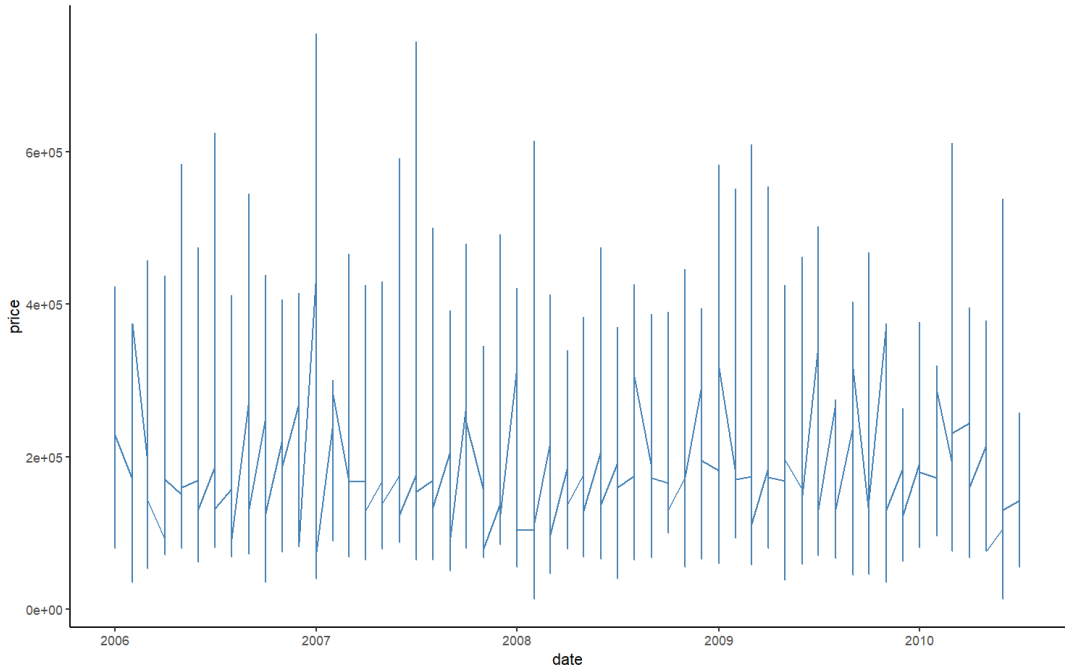
A. The function `paste0()` automatically joins elements with no separator, while `paste()` joins elements with a space by default.

B. When combining separate year, month, and day columns, using `paste(..., sep = "-")` helps create a valid date string for `as.Date()`.

C. Both `sep` and `collapse` can be used in `paste()` to control how elements and vectors are concatenated.

D. If a date is written as `"2025/10/05"`, the function `as.Date("2025/10/05")` will automatically recognize it as a valid date without specifying a format.

15. (**2 points**) In the plot below, many short vertical segments appear instead of a single smooth line for each month. Which statement best explains this behavior?



A. The vertical lines indicate the standard deviation of the value within that date.

B. There are multiple rows in the dataframe that share the same date.

C. The y variable must be transformed with `log()` before using `geom_line()`.

D. Because the x variable is a character, converting it to numeric fixes it.

Answer Key

1.D 2.B 3.C 4.C 5.D 6.D 7.B 8.A 9.C 10.A 11.C 12.A 13.B 14.D 15.B