quiz5

YANING JIN

Simulate

```
set.seed(123)
matt_pages <- sample(20:100, 100, replace = TRUE)
rol_pages <- sample(20:100, 100, replace = TRUE)
mike_pages <- sample(20:100, 100, replace = TRUE)
mean_pages <- c(60, 60)
cor_matrix <- matrix(c(1, 0.7, 0.7, 1), 2, 2)
correlated_pages <- mvrnorm(n = 100, mu = mean_pages, Sigma = cor_matrix * 15^2)
ash_pages <- correlated_pages[,1]
jacki_pages <- correlated_pages[,2]
days <- 1:100
reading_data <- data.frame(days, matt_pages, ash_pages, jacki_pages, rol_pages, mike_pages
# Test 1
cor_ash_jacki <- cor(reading_data$ash_pages, reading_data$jacki_pages)
print(paste("Correlation between Ash and Jacki's reading pages:", cor_ash_jacki))
```

```
[1] "Correlation between Ash and Jacki's reading pages: 0.694793228314227"
```

```
# Test 2
mean_pages_read <- colMeans(reading_data[,-1])
print("Mean pages read by each student:")
```

```
[1] "Mean pages read by each student:"
```

```
print(mean_pages_read)
```

```
 matt_pages   ash_pages jacki_pages   rol_pages  mike_pages
   60.26000    62.48622    61.63402    57.86000    59.97000
```

```r
# Test 3
t_test_ash_vs_matt <- t.test(reading_data$ash_pages, reading_data$matt_pages)
print("T-test result Ash vs Matt:")
```

```
[1] "T-test result Ash vs Matt:"
```

```r
print(t_test_ash_vs_matt)
```

```

    Welch Two Sample t-test

data:  reading_data$ash_pages and reading_data$matt_pages
t = 0.7949, df = 172.15, p-value = 0.4278
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.301803  7.754247
sample estimates:
mean of x mean of y
 62.48622  60.26000
```

```r
# Test 4
variance_pages_read <- apply(reading_data[,-1], 2, var)
print("Variance of pages read per day for each student:")
```

```
[1] "Variance of pages read per day for each student:"
```

```r
print(variance_pages_read)
```

```
 matt_pages   ash_pages jacki_pages   rol_pages  mike_pages
   544.1539    240.2073    215.0522    492.3236    523.4637
```

```r
# Test 5
max_pages_read <- apply(reading_data[,-1], 2, max)
print("Maximum pages read in a day by each student:")
```

```
[1] "Maximum pages read in a day by each student:"
```

```r
print(max_pages_read)
```

```
 matt_pages   ash_pages jacki_pages   rol_pages  mike_pages
  100.00000    96.49375    96.46189   100.00000   100.00000
```

Explore

```r
library(ggplot2)
set.seed(123)
days <- 100
avg_pages <- c(Matt = 50, Ash = 45, Jacki = 30, Rol = 40, Mike = 35)
matt_pages <- rnorm(days, mean=avg_pages["Matt"], sd=10)
ash_pages <- matt_pages + rnorm(days, mean=5, sd=5)
jacki_pages <- rnorm(days, mean=avg_pages["Jacki"], sd=15)
rol_pages <- rnorm(days, mean=avg_pages["Rol"], sd=10)
mike_pages <- rnorm(days, mean=avg_pages["Mike"], sd=8)
matt_pages[matt_pages < 0] <- 0
ash_pages[ash_pages < 0] <- 0
jacki_pages[jacki_pages < 0] <- 0
rol_pages[rol_pages < 0] <- 0
mike_pages[mike_pages < 0] <- 0
df_pages <- data.frame(
  Day = 1:days,
  Matt = matt_pages,
  Ash = ash_pages,
  Jacki = jacki_pages,
  Rol = rol_pages,
  Mike = mike_pages
)
ggplot(df_pages, aes(x = Day)) +
  geom_line(aes(y = Matt, color = "Matt")) +
```

```
geom_line(aes(y = Ash, color = "Ash")) +
geom_line(aes(y = Jacki, color = "Jacki")) +
geom_line(aes(y = Rol, color = "Rol")) +
geom_line(aes(y = Mike, color = "Mike")) +
labs(title = "Daily Pages Read by Each Undergraduate Over 100 Days",
     x = "Day", y = "Pages Read") +
scale_color_manual(values = c("Matt" = "blue", "Ash" = "red", "Jacki" = "green", "Rol" =
theme_minimal()
```