

# Quiz5

HengMa

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

Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
(as 'lib' is unspecified)

```
install.packages("MASS")
```

Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
(as 'lib' is unspecified)

```
library(ggplot2)  
library(MASS)
```

```
set.seed(1)
```

```
days <- 1:100  
matt_pages <- rnorm(100, mean=50, sd=10)  
rol_pages <- rnorm(100, mean=60, sd=15)  
mike_pages <- rnorm(100, mean=40, sd=5)
```

```
mean_pages <- c(45, 55)  
sigma <- matrix(c(100, 80, 80, 100), ncol=2)  
correlated_pages <- mvrnorm(n=100, mu=mean_pages, Sigma=sigma)  
ash_pages <- correlated_pages[,1]
```

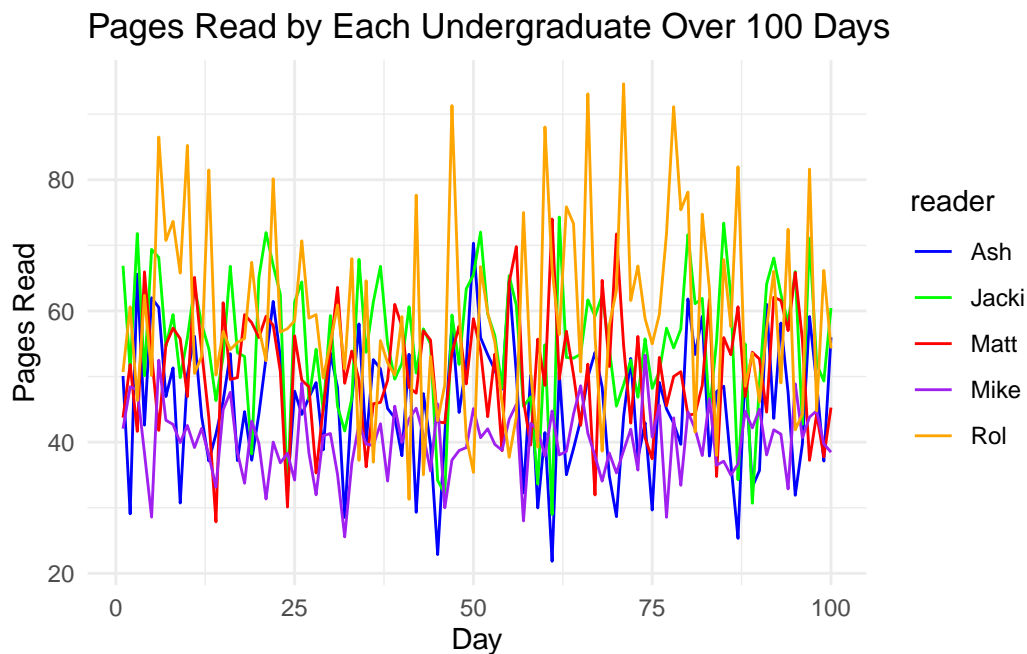
```

jacki_pages <- correlated_pages[,2]

data <- data.frame(day = rep(days, 5),
                  pages = c(matt_pages, rol_pages, mike_pages, ash_pages, jacki_pages),
                  reader = factor(rep(c("Matt", "Rol", "Mike", "Ash", "Jacki"), each=100))

ggplot(data, aes(x=day, y=pages, color=reader)) +
  geom_line() +
  ggtitle("Pages Read by Each Undergraduate Over 100 Days") +
  xlab("Day") +
  ylab("Pages Read") +
  theme_minimal() +
  scale_color_manual(values=c("blue", "green", "red", "purple", "orange"))

```



## Explaintion

To simulate the reading habits of five undergraduates—Matt, Ash, Jacki, Rol, and Mike—over 100 days, with each individual's daily page counts independent from the others, we use R to generate random data reflecting their unique reading patterns. By assigning each person a mean and standard deviation for their daily reading, we create a diverse set of

reading behaviors, without any correlation, even between Ash and Jacki who were previously considered as a couple. This data is then visualized using ggplot2, producing a line graph that color-codes each undergraduate's reading journey across the days, clearly showcasing their individual reading habits in a straightforward, comparative manner. This approach provides a simple yet effective way to analyze and display the independent reading trends of each student over the specified period.

## **Five Tests**

### **Descriptive Statistics Test**

Calculate the mean, median, standard deviation, minimum, and maximum number of pages read by each undergraduate.

### **One-Way ANOVA (Analysis of Variance)**

To determine if there are statistically significant differences in the average number of pages read between the undergraduates

### **Kruskal-Wallis Test**

If the data do not meet the assumptions of normality required for ANOVA, the Kruskal-Wallis test, a non-parametric alternative to ANOVA, can be used.

### **Pairwise Comparison Test**

Following ANOVA or Kruskal-Wallis, if we find significant differences, pairwise comparison tests like Tukey's HSD (for ANOVA) or Dunn's test (for Kruskal-Wallis) can be conducted.

### **Time Series Analysis**

Since the data is over a period of 100 days, conducting a time series analysis on the number of pages read daily by each undergraduate could reveal trends, cycles, or patterns in their reading habits over time.