quiz5

2024-02-06

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

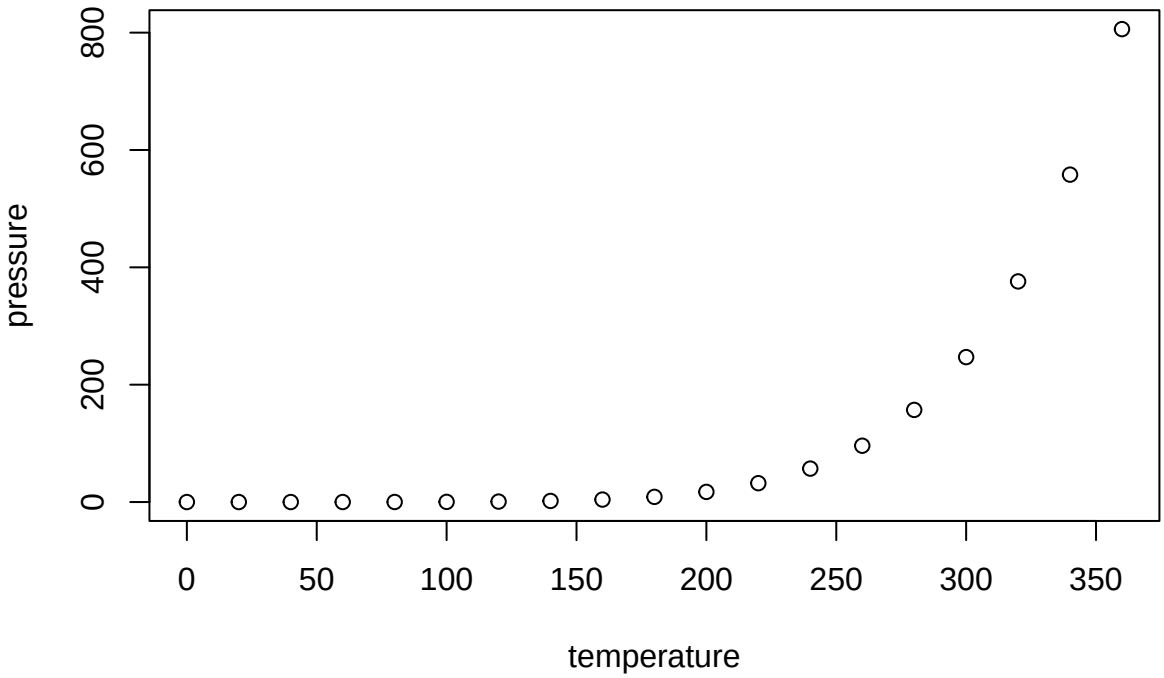
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```r
# Setting seed for reproducibility
set.seed(42)

# Number of days
days <- 100

# Simulate reading pages for each undergraduate, assuming independence
matt <- rnorm(days, mean=30, sd=10)
ash <- rnorm(days, mean=50, sd=15)
jacki <- rnorm(days, mean=20, sd=5)
rol <- rnorm(days, mean=45, sd=10)
mike <- rnorm(days, mean=40, sd=20)

# Ensuring no negative page counts
matt <- ifelse(matt < 0, 0, matt)
ash <- ifelse(ash < 0, 0, ash)
jacki <- ifelse(jacki < 0, 0, jacki)
rol <- ifelse(rol < 0, 0, rol)
mike <- ifelse(mike < 0, 0, mike)

# Creating a data frame
data <- data.frame(Day=1:days, Matt=matt, Ash=ash, Jacki=jacki, Rol=rol, Mike=mike)

# Performing five tests based on the simulated data

# 1. Test for normality for Matt's reading pattern
shapiro.test(data$Matt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$Matt
## W = 0.98122, p-value = 0.1654
```

```r
# 2. Test for normality for Ash's reading pattern
shapiro.test(data$Ash)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$Ash
## W = 0.98732, p-value = 0.4595
```

```r
# 3. Correlation test between Matt and Ash (expecting low correlation due to independence)
cor.test(data$Matt, data$Ash)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$Matt and data$Ash
## t = 0.30981, df = 98, p-value = 0.7574
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1661592  0.2263075
## sample estimates:
```

```
##        cor
## 0.03127984
```

```
# 4. Variance analysis - ANOVA to compare means across all undergraduates
anova(lm(cbind(Matt, Ash, Jacki, Rol, Mike) ~ Day, data=data))
```

```
## Analysis of Variance Table
##
##             Df  Pillai approx F num Df den Df Pr(>F)
## (Intercept)  1 0.98518  1249.39      5     94 <2e-16 ***
## Day          1 0.02739     0.53      5     94 0.7535
## Residuals   98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 5. Pairwise comparison of mean pages read per day between all undergraduates using Tukey's HSD test

install.packages("reshape2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(reshape2)
data_long <- melt(data, id.vars = "Day", variable.name = "Undergraduate", value.name = "Pages")

# ANOVA to compare means across all undergraduates
model <- aov(Pages ~ Undergraduate, data=data_long)
summary(model)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## Undergraduate   4  54029   13507   84.74 <2e-16 ***
## Residuals     495  78899     159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Tukey's HSD test for pairwise comparison between all undergraduates
TukeyHSD(model)
```

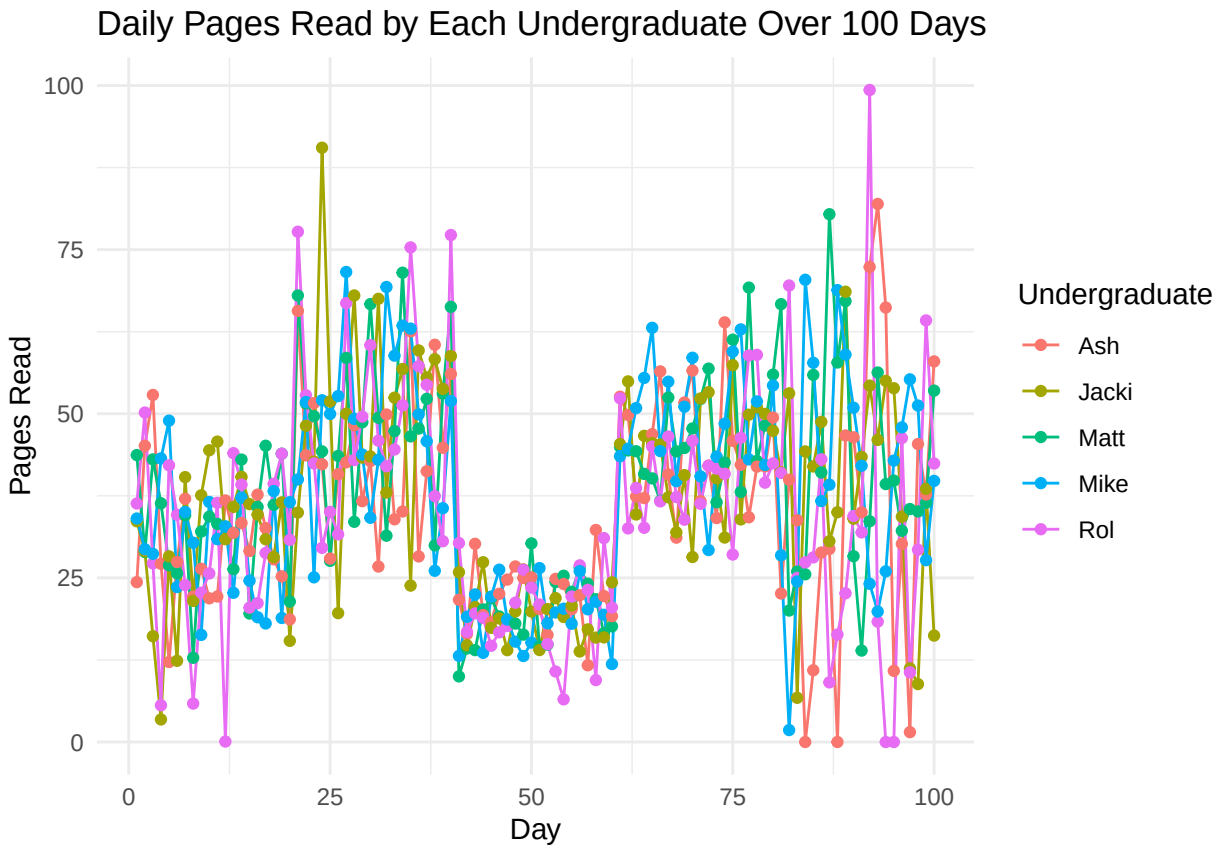
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Pages ~ Undergraduate, data = data_long)
##
## $Undergraduate
##                  diff        lwr        upr     p adj
## Ash-Matt    18.362596  13.474281  23.250911 0.0000000
## Jacki-Matt -10.376989 -15.265304  -5.488674 0.0000001
## Rol-Matt    15.004216  10.115901  19.892532 0.0000000
## Mike-Matt    7.490227   2.601912  12.378542 0.0003112
## Jacki-Ash  -28.739585 -33.627900 -23.851270 0.0000000
## Rol-Ash     -3.358380  -8.246695   1.529935 0.3289468
## Mike-Ash   -10.872369 -15.760684  -5.984054 0.0000000
## Rol-Jacki   25.381206  20.492890  30.269521 0.0000000
## Mike-Jacki  17.867216  12.978901  22.755531 0.0000000
## Mike-Rol    -7.513990 -12.402305  -2.625674 0.0002942
```

```
library(ggplot2)
```

```r
set.seed(42)
days <- 100
data <- data.frame(
  Day = rep(1:days, each = 5),
  Undergraduate = rep(c("Matt", "Ash", "Jacki", "Rol", "Mike"), times = days),
  Pages = c(
    rnorm(days, 30, 10),
    rnorm(days, 50, 15),
    rnorm(days, 20, 5),
    rnorm(days, 45, 10),
    rnorm(days, 40, 20)
  )
)
data$Pages[data$Pages < 0] <- 0

# Plotting with ggplot2
p <- ggplot(data, aes(x = Day, y = Pages, color = Undergraduate)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Daily Pages Read by Each Undergraduate Over 100 Days",
       x = "Day",
       y = "Pages Read",
       color = "Undergraduate")

# Display the plot
print(p)
```


Daily Pages Read by Each Undergraduate Over 100 Days
Pages Read
100
75
50
25
0
0
25
50
75
100
Day
Undergraduate
Ash
Jacki
Matt
Mike
Rol