PART I: DATA COLLECTION AND DATA PROCESSING

Problem 1

A. The purpose of the data collection. Hypothesis: Increased daily screen time is associated with decreased quality of mental health among adults.

The primary purpose of this data collection is to empirically examine the relationship between daily screen time and mental health outcomes in adults. In the digital age, adults are increasingly engaging with various forms of screen media, including smartphones, computers, tablets, and televisions. While these devices provide valuable opportunities for communication, work, and entertainment, there is growing concern about the potential negative impacts of prolonged screen exposure on mental well-being (Coyne et.al, 2020). This study aims to collect and analyze data reflecting individuals' screen activity patterns and correlate these findings with self-reported measures of mental health.

B. The role of Informed Consent Form. It ensures that participants are fully aware of the study's nature, purpose, procedures, risks, and benefits. It explains how the data will be used to test the hypothesis, and assure participants of their privacy and the confidentiality of their data. Obtain their consent to use the data for research purposes.

C. The data collection plan.

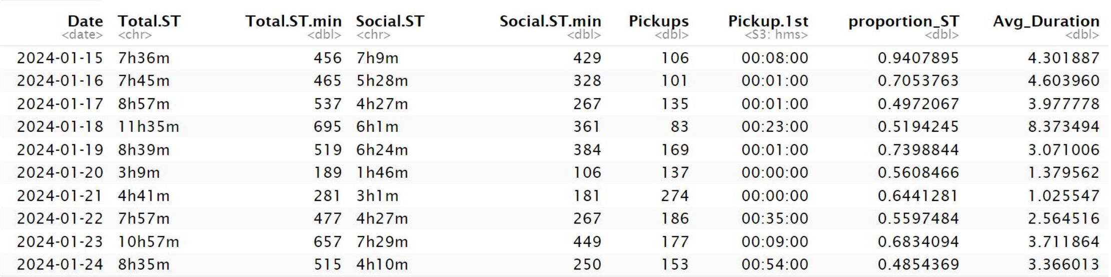
When Collected: Data is collected continuously from the start of the study (January 15, 2024) until the data freeze date of Friday, January 26, 2024.

Variables Collected: Total daily screen time, daily social media screen time, daily number of device pickups, the first time that pick up the phone each day.

Source of Data: Data is collected from the digital well-being settings of participants' mobile devices or through a dedicated app designed to track device usage.

Volume of Data: 2024-01-15 to 2024-01-26

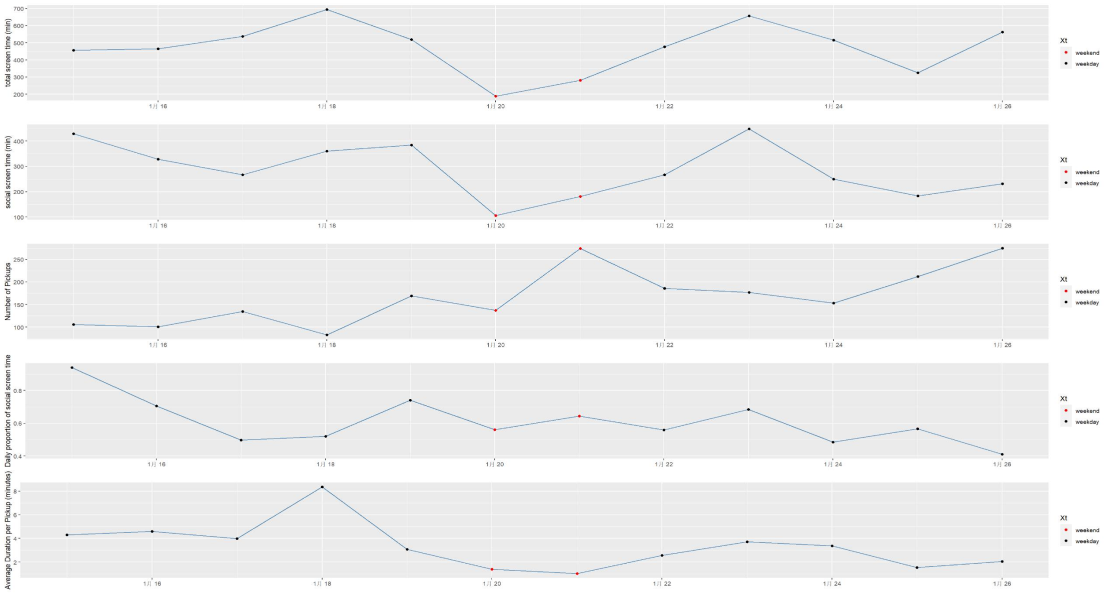
D. Create two new variables.



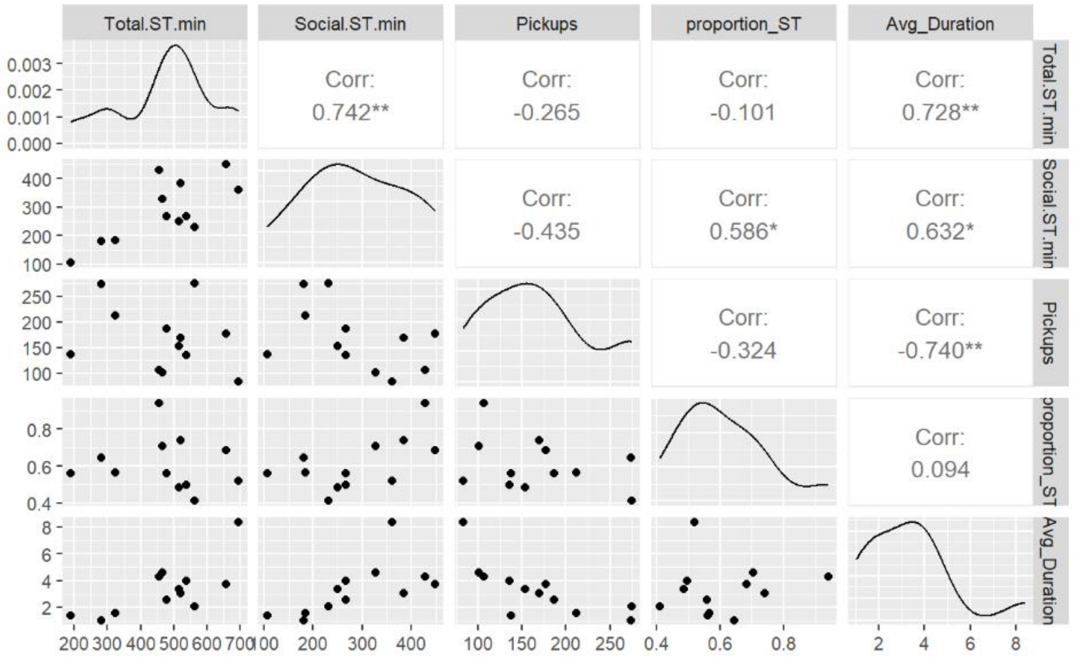
Date <date>	Total.ST <chr>	Total.ST.min <dbl>	Social.ST <chr>	Social.ST.min <dbl>	Pickups <dbl>	Pickup.1st <S3: hms>	proportion_ST <dbl>	Avg_Duration <dbl>
2024-01-15	7h36m	456	7h9m	429	106	00:08:00	0.9407895	4.301887
2024-01-16	7h45m	465	5h28m	328	101	00:01:00	0.7053763	4.603960
2024-01-17	8h57m	537	4h27m	267	135	00:01:00	0.4972067	3.977778
2024-01-18	11h35m	695	6h1m	361	83	00:23:00	0.5194245	8.373494
2024-01-19	8h39m	519	6h24m	384	169	00:01:00	0.7398844	3.071006
2024-01-20	3h9m	189	1h46m	106	137	00:00:00	0.5608466	1.379562
2024-01-21	4h41m	281	3h1m	181	274	00:00:00	0.6441281	1.025547
2024-01-22	7h57m	477	4h27m	267	186	00:35:00	0.5597484	2.564516
2024-01-23	10h57m	657	7h29m	449	177	00:09:00	0.6834094	3.711864
2024-01-24	8h35m	515	4h10m	250	153	00:54:00	0.4854369	3.366013

Problem 2

A. Time series plots. The patterns of the total screen time and the social screen time are quite similar. The data collected during weekend days are colored in red in order to visualize any differences of mobile device use between weekdays and weekends.



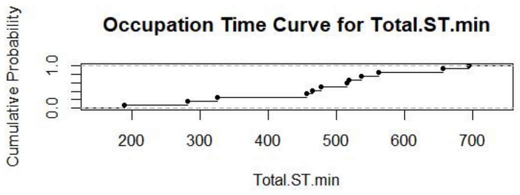
B. Pairwise scatterplots. There is a strong positive correlation between total screen time and social screen time, with the Pearson correlation equal to 0.742. The Average Duration per Pickup has a significant negative correlation (-0.740) with the Proportion of Social Screen Time, indicating that on days with a higher proportion of social screen time, the duration per pickup tends to be shorter. Number of Pickups appears to have a more uniform distribution, with a slight skew towards fewer pickups. Proportion of Social Screen Time (proportion_ST) suggests a skewed distribution towards the lower proportion values, indicating that social screen time typically constitutes a smaller proportion of the total screen time.

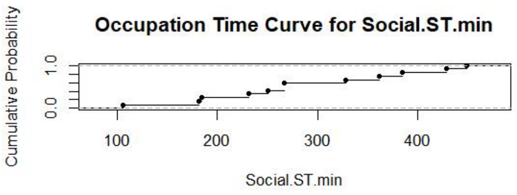


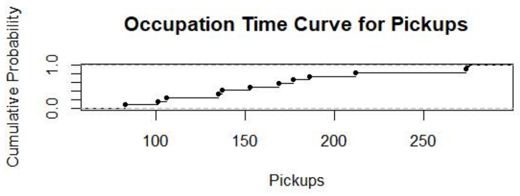
C. Occupation time curve. Total Screen Time and Social Screen Time: Both have occupation time curves with steps, indicating that certain screen time values are more common than others. Number of Pickups: The occupation curve is relatively smooth, which corresponds to a more uniform distribution of pickup counts.

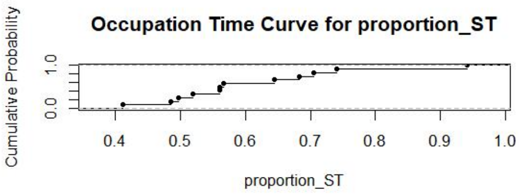
Proportion of Social Screen Time: The curve rises steeply initially and then levels off, which aligns with the skewed distribution toward lower values.

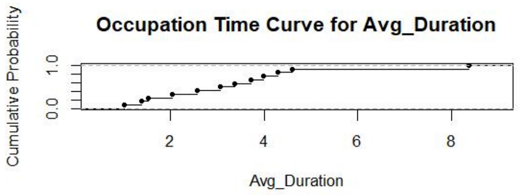
Average Duration per Pickup: The occupation curve for average duration per pickup shows a steep initial rise, flattening out as the duration increases, which is consistent with the skewness towards lower average durations.



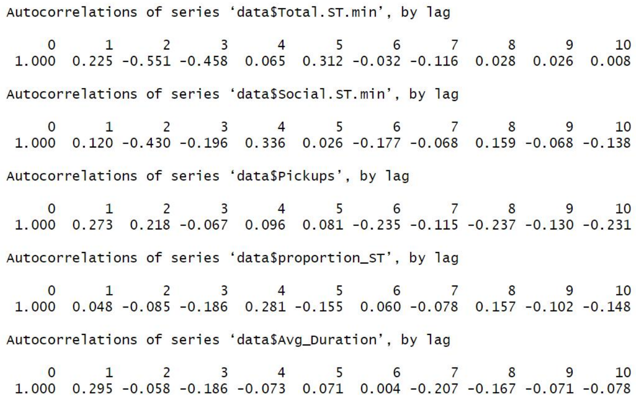








D. ACF



```
Autocorrelations of series ‘data$Total.ST.min’, by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.225 -0.551 -0.458  0.065  0.312 -0.032 -0.116  0.028  0.026  0.008 

Autocorrelations of series ‘data$Social.ST.min’, by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.120 -0.430 -0.196  0.336  0.026 -0.177 -0.068  0.159 -0.068 -0.138 

Autocorrelations of series ‘data$Pickups’, by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.273  0.218 -0.067  0.096  0.081 -0.235 -0.115 -0.237 -0.130 -0.231 

Autocorrelations of series ‘data$proportion_ST’, by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.048 -0.085 -0.186  0.281 -0.155  0.060 -0.078  0.157 -0.102 -0.148 

Autocorrelations of series ‘data$Avg_Duration’, by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.295 -0.058 -0.186 -0.073  0.071  0.004 -0.207 -0.167 -0.071 -0.078 
```

For data$Total.ST.min, the lag-2 autocorrelation is notably negative (-0.551). This suggests that there may be a negative correlation between a day's total screen time and the previous day's total screen time.

For data$Social.ST.min, there is a negative lag-2 autocorrelation (-0.430) that stands out. This indicates that days with higher social screen time may be followed by days with lower social screen time, or vice versa.

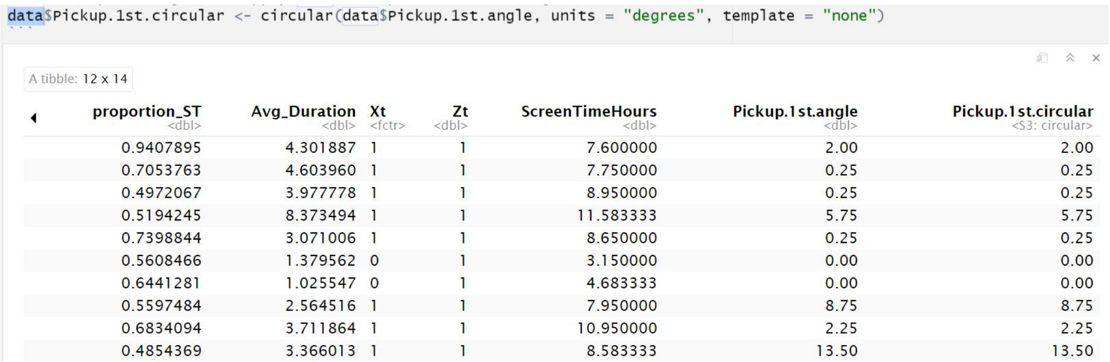
For data$Pickups, the lag-1 autocorrelation is positive (0.273), which may suggest that the number of pickups is somewhat consistent from one day to the next.

For data$proportion_ST, there is a positive lag-4 autocorrelation (0.281) that could be of interest, indicating a potential weekly pattern since the data likely covers consecutive days.

For data$Avg_Duration, there are no autocorrelation values close to 1 or -1, suggesting no significant autocorrelation at the given lags.

Problem 3

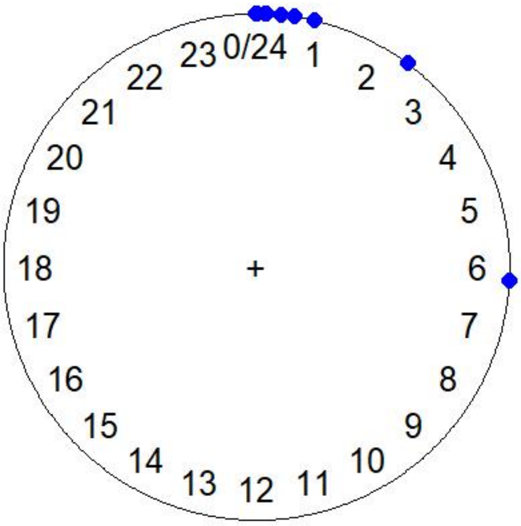
A. Transform the time.



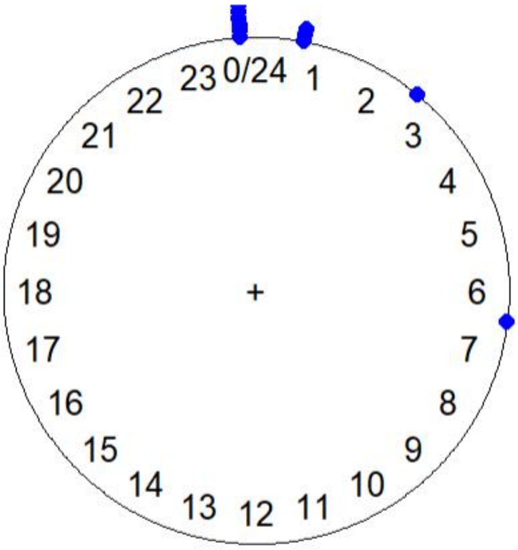
```
data$Pickup.1st.circular <- circular(data$Pickup.1st.angle, units = "degrees", template = "none")
```

```
A tibble: 12 x 14
   proportion_ST  Avg_Duration  Xt      Zt  ScreenTimeHours  Pickup.1st.angle  Pickup.1st.circular
           <dbl>         <dbl>  <fctr> <dbl>          <dbl>             <dbl>       <S3: circular>
       0.9407895      4.301887  1        1        7.600000              2.00                 2.00
       0.7053763      4.603960  1        1        7.750000              0.25                 0.25
       0.4972067      3.977778  1        1        8.950000              0.25                 0.25
       0.5194245      8.373494  1        1       11.583333              5.75                 5.75
       0.7398844      3.071006  1        1        8.650000              0.25                 0.25
       0.5608466      1.379562  0        1        3.150000              0.00                 0.00
       0.6441281      1.025547  0        1        4.683333              0.00                 0.00
       0.5597484      2.564516  1        1        7.950000              8.75                 8.75
       0.6834094      3.711864  1        1       10.950000              2.25                 2.25
       0.4854369      3.366013  1        1        8.583333             13.50                13.50
```

B. Circular scatter plot of 1st pickup time.



C. Histogram of 1st pickup time.

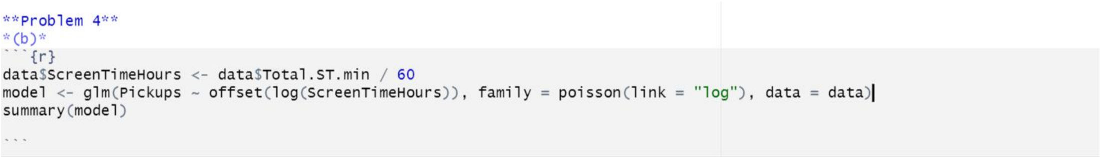


PART II: DATA ANALYSIS

Problem 4

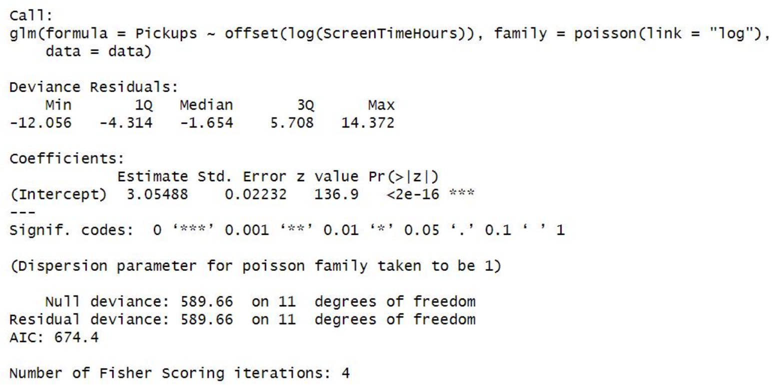
A. Explain why the factor St is needed in the Poisson distribution above. The Poisson distribution assumes that the number of events (pickups) in non-overlapping intervals is independent, and the expected number of events is proportional to the length of the interval (in this case, the screen time). Therefore, if a person has more screen time on a particular day, the opportunity for pickups increases proportionally.By including the parameter λ can be interpreted as the expected rate of pickups per hour of screen time, rather than the total expected number of pickups per day. This standardization allows for meaningful comparison of λ across different days or individuals with varying screen time duration.

B.



~~~
**Problem 4**
*(b)*
```{r}
data$ScreenTimeHours <- data$Total.ST.min / 60
model <- glm(Pickups ~ offset(log(ScreenTimeHours)), family = poisson(link = "log"), data = data)
summary(model)

~~~
```



```
Call:
glm(formula = Pickups ~ offset(log(ScreenTimeHours)), family = poisson(link = "log"),
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-12.056   -4.314   -1.654    5.708   14.372

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.05488    0.02232   136.9   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

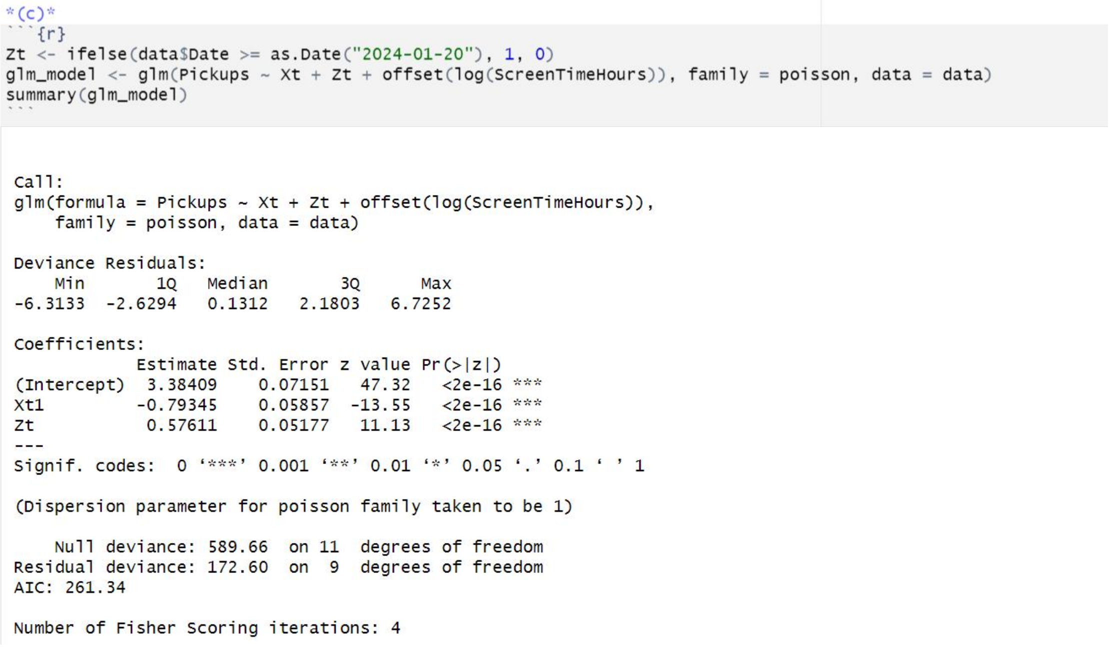
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 589.66  on 11  degrees of freedom
Residual deviance: 589.66  on 11  degrees of freedom
AIC: 674.4

Number of Fisher Scoring iterations: 4
```

C.

In this part, I used the time 1.20, cause my flight is canceled and just arrived in UMCIH on January 20.



```
*(c)*
```{r}
Zt <- ifelse(data$Date >= as.Date("2024-01-20"), 1, 0)
glm_model <- glm(Pickups ~ Xt + Zt + offset(log(ScreenTimeHours)), family = poisson, data = data)
summary(glm_model)
```

Call:
glm(formula = Pickups ~ Xt + Zt + offset(log(ScreenTimeHours)),
    family = poisson, data = data)

Deviance Residuals:
    Min   1Q   Median   3Q   Max
-6.3133 -2.6294   0.1312   2.1803   6.7252

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 3.38409   0.07151   47.32   <2e-16 ***
Xt1   -0.79345   0.05857 -13.55   <2e-16 ***
Zt   0.57611   0.05177   11.13   <2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 589.66 on 11 degrees of freedom
Residual deviance: 172.60 on 9 degrees of freedom
AIC: 261.34

Number of Fisher Scoring iterations: 4
```
```

C.1

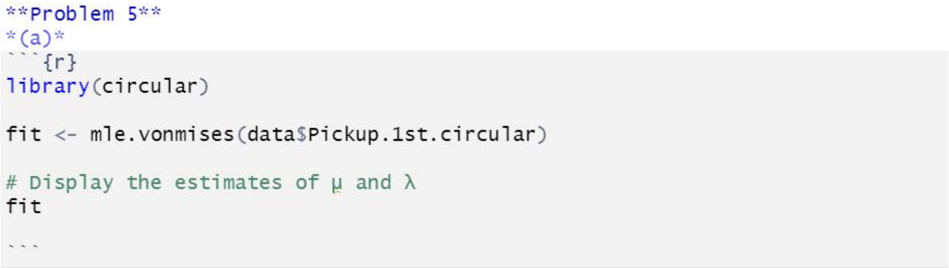
Looking at the coefficient for Xt , we see that the estimate is -0.79345 with a very small p-value (less than 2e-16), which is highly significant. The negative sign of the coefficient indicates that the rate of pickupsλ(t) on weekends is lower than on weekdays. Since the p-value is much less than the significance level (α=0.05), we have strong evidence to reject the null hypothesis that there is no difference in the behavior of daily pickups between weekdays and weekends. Therefore, there is data evidence for significantly different behavior of daily pickups between weekdays and weekends.

C.2

Regarding the coefficient for Zt, we have an estimate of 0.57611 with a p-value also less than 2e-16, which is highly significant. The positive sign indicates that the rate of pickups is higher during or after the start of the winter semester compared to before. Thus, there is data evidence for a significant change in the behavior of daily pickups after the winter semester began (the date just for me).

Problem 5

A.

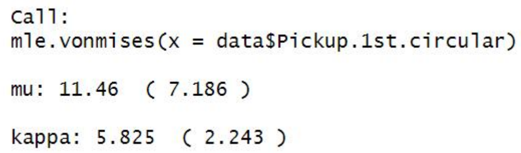


~~~
**Problem 5**
*(a)*
```{r}
library(circular)

fit <- mle.vonmises(data$Pickup.1st.circular)

# Display the estimates of μ and λ
fit

```
~~~

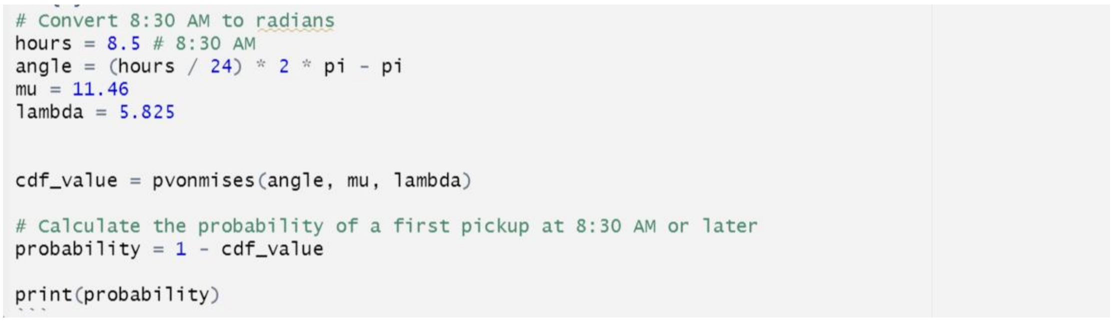


```
Call:
mle.vonmises(x = data$Pickup.1st.circular)

mu: 11.46   ( 7.186 )

kappa: 5.825   ( 2.243 )
```

B.



```
# Convert 8:30 AM to radians
hours = 8.5 # 8:30 AM
angle = (hours / 24) * 2 * pi - pi
mu = 11.46
lambda = 5.825


cdf_value = pvonmises(angle, mu, lambda)

# Calculate the probability of a first pickup at 8:30 AM or later
probability = 1 - cdf_value

print(probability)
```

```
## [1] 0.3273545
```

Reference

[1] Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in human behavior*, *104*, 106160.

Code:

Problem 1 *(d)*

```r
library(ggplot2)
library(lubridate)
library(dplyr)
library(tidyr)
library(readr)
library(gridExtra)
data <- read_csv("D:/620/data.csv", show_col_types = FALSE)
data$Date <- as.Date(data$Date, format="%Y/%m/%d")
data$proportion_ST <- data$Social.ST.min / data$Total.ST.min
data$Avg_Duration <- data$Total.ST.min / data$Pickups
Xt <- ifelse(weekdays(data$Date) %in% c("星期六", "星期日"), 0, 1)
Xt <- as.factor(Xt)
# Cause my R Studio only supports Chinese. "星期六"means Saturday,"星期日" means Sunday.
```

Problem 2 *(a)*

```r
total <- ggplot(data,aes(x = Date,y = Total.ST.min,color = Xt))+
  geom_line(color = "steelblue")+
  geom_point()+
  xlab("")+ylab("total screen time (min)")+
  scale_color_manual(labels=c("weekend","weekday"),values=c("red","black"))

social <- ggplot(data,aes(x = Date,y = Social.ST.min, color = Xt))+
  geom_line(color = "steelblue")+
  geom_point()+
  xlab("")+ylab("social screen time (min)")+
  scale_color_manual(labels=c("weekend","weekday"),values=c("red","black"))

Pickups <- ggplot(data,aes(x = Date,y = Pickups, color = Xt))+
  geom_line(color = "steelblue")+
  geom_point()+
  xlab("")+ylab("Number of Pickups")+
  scale_color_manual(labels=c("weekend","weekday"),values=c("red","black"))

proportion_st <- ggplot(data,aes(x = Date,y = proportion_ST, color = Xt))+
  geom_line(color = "steelblue")+
  geom_point()+
  xlab("")+ylab("Daily proportion of social screen time")+
  scale_color_manual(labels=c("weekend","weekday"),values=c("red","black"))

avg_duration <- ggplot(data,aes(x = Date,y = Avg_Duration, color = Xt))+
  geom_line(color = "steelblue")+
  geom_point()+
  xlab("")+ylab("Average Duration per Pickup (minutes)")+
  scale_color_manual(labels=c("weekend","weekday"),values=c("red","black"))
```

```
grid.arrange(total, social, Pickups, proportion_st, avg_duration, nrow = 5)
```

(b)

```
library(GGally)
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
numeric_data <- data %>% select(Total.ST.min, Social.ST.min, Pickups, proportion_ST, Avg_D
uration)

ggpairs(numeric_data)
```

(c)

```
plot_ecdf <- function(data, column) {
  ecdf_data <- ecdf(data[[column]])
  plot(ecdf_data, main = paste("Occupation Time Curve for", column), xlab = column, yla
b = "Cumulative Probability")
}

# Apply the function to each of the five variables
par(mfrow = c(3, 2))  # Organize the plots into a 3x2 grid
plot_ecdf(data, "Total.ST.min")
plot_ecdf(data, "Social.ST.min")
plot_ecdf(data, "Pickups")
# Assuming you also have a variable for 'Proportion of Social Screen Time' and 'Duration
 per Use'
plot_ecdf(data, "proportion_ST")
plot_ecdf(data, "Avg_Duration")
```

(d)

```
acf(data$Total.ST.min, plot = FALSE)
acf(data$Social.ST.min, plot = FALSE)
acf(data$Pickups, plot = FALSE)
acf(data$proportion_ST, plot = FALSE)
acf(data$Avg_Duration, plot = FALSE)
```

Problem 3 *(a)*

```
library(circular)
##
## 载入程辑包：'circular'
## The following objects are masked from 'package:stats':
##
##     sd, var
data$Pickup.1st <- as.POSIXct(data$Pickup.1st, format = "%H:%M:%S")

# Function to convert time to angle
time_to_angle <- function(time) {
  # Extract hours and minutes
  hrs <- as.numeric(format(time, "%H"))
  mins <- as.numeric(format(time, "%M"))

  # Calculate the angle (0 degrees at midnight, 360 degrees at next midnight)
  angle <- (hrs * 60 + mins) / (24 * 60) * 360
  return(angle)
}

# Apply the function to the time column to create a new angle column
data$Pickup.1st.angle <- sapply(data$Pickup.1st, time_to_angle)
data$Pickup.1st.circular <- circular(data$Pickup.1st.angle, units = "degrees", template = "none")
```

(b)

```
pickup_circular <- circular(data$Pickup.1st.angle, units = "degrees",template='clock24' )

plot(pickup_circular, col="blue")
```

(c)

```
plot(pickup_circular, stack=TRUE, bins=25, col="blue")
```

Problem 4 *(b)*

```
data$ScreenTimeHours <- data$Total.ST.min / 60
model <- glm(Pickups ~ offset(log(ScreenTimeHours)), family = poisson(link = "log"), data = data)
summary(model)
```

(c)

```
Zt <- ifelse(data$Date >= as.Date("2024-01-20"), 1, 0)
glm_model <- glm(Pickups ~ Xt + Zt + offset(log(ScreenTimeHours)), family = poisson, dat
a = data)
summary(glm_model)
```

Problem 5 *(a)*

```
library(circular)

fit <- mle.vonmises(data$Pickup.1st.circular)

# Display the estimates of μ and λ
fit
```

(b)

```
# Convert 8:30 AM to radians
hours = 8.5 # 8:30 AM
angle = (hours / 24) * 2 * pi - pi
mu = 11.46
lambda = 5.825

cdf_value = pvonmises(angle, mu, lambda)
# Calculate the probability of a first pickup at 8:30 AM or later
probability = 1 - cdf_value

print(probability)
## [1] 0.3273545
```