BIOSTAT 620

Homework 1

Neyan Deng

Problem1

(a)

The purpose of the data collection is to investigate the relationship between daily mood traits and the usage of smartphone. Informed by Beierle et al. (2020), I hypothesize that specific mood traits, such as levels of extraversion or neuroticism, may correlate with the frequency and duration of smartphone use.

(b)

The Informed Consent From (ICF) have several roles. First, it informs the participants about the purpose, expectation, steps, and time of duration of the study. It informs participants understanding of their roles and discloses any potential risks. It also informs participants about whether any benefits they could gained. Meanwhile, it protects participants their autonomy so that they can make their own decision about whether getting involved or not. Moreover, it explains how the data would be collected from participants, and how those data would be used and protected. Besides, the ICF emphasizes that the participation in the study is voluntary, and informs participants are free to withdraw from the study at any time without reasons with no penalty. It provides the contact information of the relevant researchers. Additionally, the ICF serves as a legal instrument to make sure that the study complies with ethical standards and legal requirements.

(c)

The data was collected from Jan. 3rd to Jan. 26th, 2024, collected from my own cell phone. Six variables are involved, including total screen time in HH-MM format (Total.ST), total screen time in minute (Total.ST.min), social app screen time in HH-MM format (Social.ST), social screen time in minute (Social.ST.min), total number of times that I picked up the phone (Pickups), and the time of the first pickup (Pickup.1st).Therefore, the dataset initially includes six variables and 24 observations (denoted as date) before the data freeze.

(d)

Date	Total.ST	Total.ST.min	Social.ST	Social.ST.min	Pickups	Pickup.1st	Proportion.Social.ST	Duration.Per.Use
1/3/2024	6h17m	377	4h3m	243	77	00:01:00	0.6446	4.8961
1/4/2024	4h12m	252	2h19m	139	97	07:16:00	0.5516	2.5979
1/5/2024	8h0m	480	5h52m	352	176	07:21:00	0.7333	2.7273
1/6/2024	7h44m	464	4h1m	241	137	00:08:00	0.5194	3.3869
1/7/2024	3h24m	204	1h46m	106	189	00:03:00	0.5196	1.0794
1/8/2024	10h27m	627	4h28m	268	87	00:06:00	0.4274	7.2069
1/9/2024	8h3m	483	5h2m	302	123	00:09:00	0.6253	3.9268
1/10/2024	5h28m	328	1h33m	93	76	00:06:00	0.2835	4.3158
1/11/2024	5h50m	350	2h35m	155	164	00:03:00	0.4429	2.1341
1/12/2024	8h5m	485	3h2m	182	155	00:41:00	0.3753	3.1290
1/13/2024	7h59m	479	3h43m	223	102	00:03:00	0.4656	4.6961
1/14/2024	12h40m	760	3h47m	227	69	00:17:00	0.2987	11.0145
1/15/2024	11h34m	694	4h8m	248	108	00:06:00	0.3573	6.4259
1/16/2024	9h20m	560	2h11m	131	90	00:30:00	0.2339	6.2222
1/17/2024	9h28m	568	3h0m	180	98	00:16:00	0.3169	5.7959
1/18/2024	9h49m	589	3h45m	225	115	00:17:00	0.3820	5.1217
1/19/2024	8h25m	505	2h33m	153	108	00:05:00	0.3030	4.6759
1/20/2024	7h13m	433	2h17m	137	106	00:15:00	0.3164	4.0849
1/21/2024	8h33m	513	5h5m	305	105	00:03:00	0.5945	4.8857
1/22/2024	6h22m	382	2h1m	121	225	02:23:00	0.3168	1.6978
1/23/2024	8h41m	521	3h49m	229	235	03:33:00	0.4395	2.2170
1/24/2024	7h4m	424	3h7m	187	176	02:24:00	0.4410	2.4091
1/25/2024	6h51m	411	2h46m	166	137	04:25:00	0.4039	3.0000
1/26/2024	9h52m	592	3h29m	209	235	01:45:00	0.3530	2.5191

Problem2

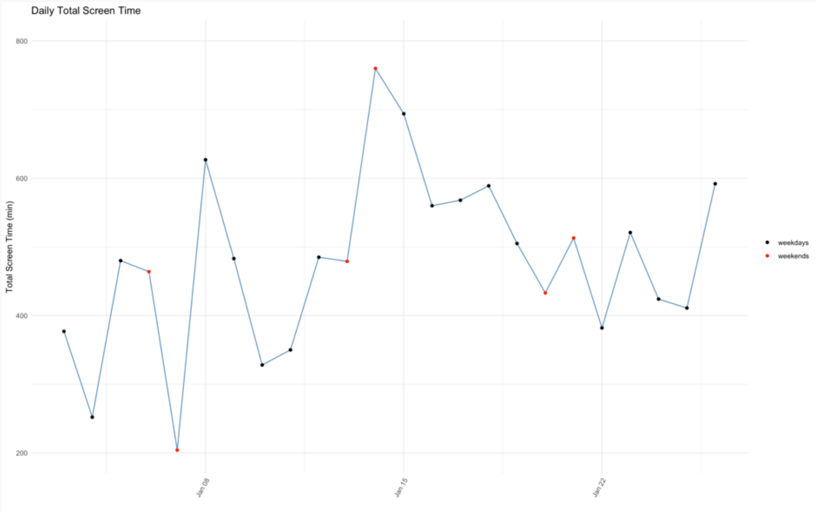
```
library(ggplot2)
library(dplyr)
library(lubridate)
library(readr)
screen_data = read_csv("~/Desktop/2024Spring/BIOSTAT620/BIOSTAT 620 Screen Data.csv")

hm_to_min <- function(hm) {
 unlist(lapply(hm, function(x) {
  splt <- strsplit(x, "h")[[1]]
  hr <- as.numeric(splt[1])
  mn <- as.numeric(strsplit(splt[2], "m")[[1]][1])
  return(60 * hr + mn)
 }))
}

screen_data <- screen_data %>%
 mutate(
  Total.ST.min.true = hm_to_min(Total.ST),
  Social.ST.min.true = hm_to_min(Social.ST),
  Total.ST.match = Total.ST.min.true == Total.ST.min,
  Social.ST.match = Social.ST.min.true == Social.ST.min
 ) %>%
 relocate(
  Date,
  Total.ST,
  Total.ST.min,
  Total.ST.min.true,
  Total.ST.match,
  Social.ST,
  Social.ST.min,
  Social.ST.min.true,
  Social.ST.match
 )
screen_data$Date <- as.Date(screen_data$Date, format = "%m/%d/%Y")
screen_data$weekday = weekdays(screen_data$Date , abbreviate = T)
screen_data = screen_data %>% mutate(if_weekend = weekday %in% c('Sun', 'Sat'))
```

(a)

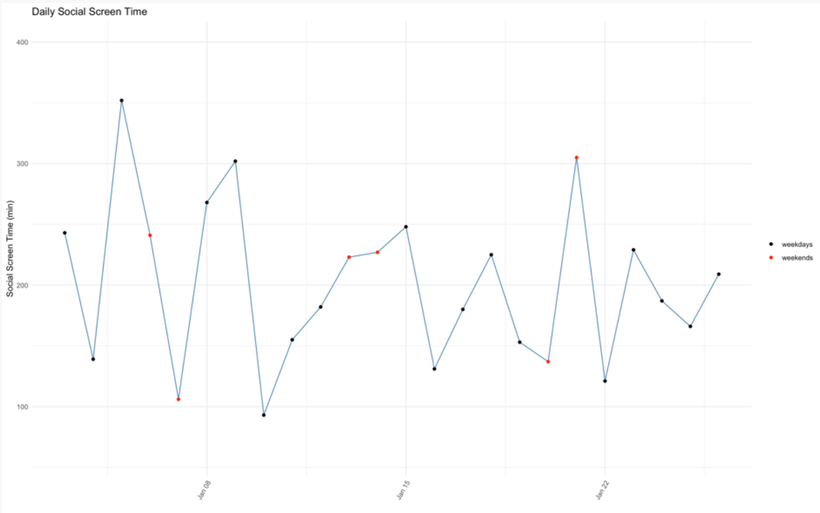
Generally, the pattern of total screen time and social screen time seems to have a similar trend, but Jan.14th is a special case caused by one major event of my personal life. Due to the fever that I caught, I lay on bed for almost the whole day, reading fictions through my phone, which makes my usage of screen extremely high, with relatively low pickup-times and extremely high duration per use. The proportion of social screen time is relatively higher before school started.

```
total = ggplot(screen_data, aes(x=Date, y=Total.ST.min.true, color=if_weekend)) +
 geom_line(color="steelblue") +
 geom_point() +
 xlab("") + ylab("Total Screen Time (min)") +
 ggtitle("Daily Total Screen Time")+
 ylim(200, 800) +
 scale_color_manual(labels=c("weekdays", "weekends"), values=c("black", "red")) +
 theme_minimal() +
```

```
 theme(axis.text.x=element_text(angle=60, hjust=1), legend.title=element_blank())
print(total)
```

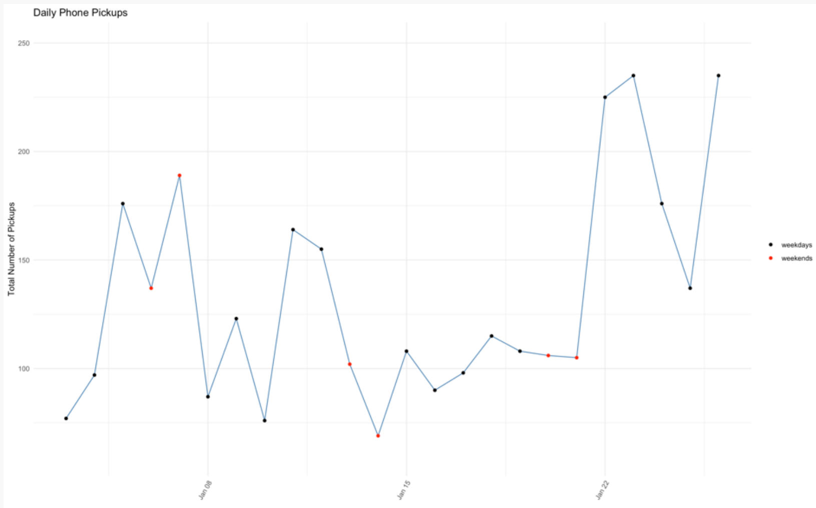



```
social = ggplot(screen_data, aes(x=Date, y=Social.ST.min.true, color=if_weekend)) +
 geom_line(color="steelblue") +
 geom_point() +
 xlab("") + ylab("Social Screen Time (min)") +
 ggtitle("Daily Social Screen Time")+
 ylim(60, 400) +
 scale_color_manual(labels=c("weekdays", "weekends"), values=c("black", "red")) +
 theme_minimal() +
 theme(axis.text.x=element_text(angle=60, hjust=1), legend.title=element_blank())
print(social)
```

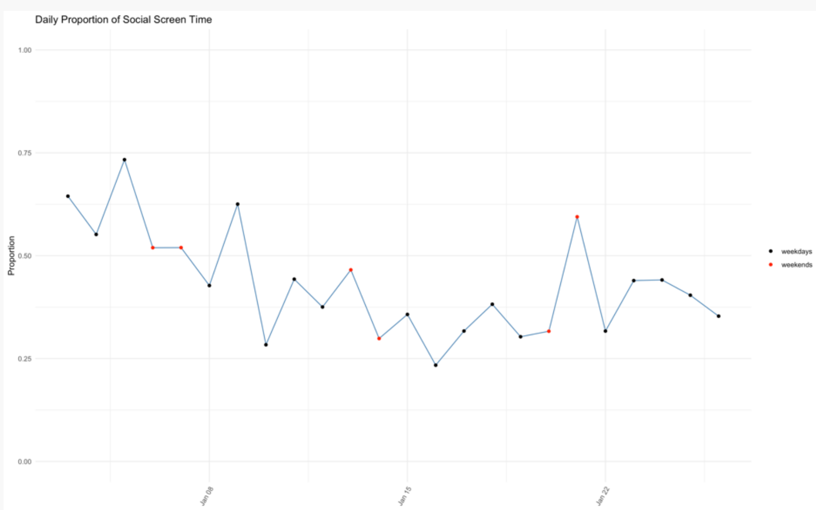



```
pickups = ggplot(screen_data, aes(x=Date, y=Pickups, color=if_weekend)) +
 geom_line(color="steelblue") +
 geom_point() +
 xlab("") + ylab("Total Number of Pickups") +
 ggtitle("Daily Phone Pickups")+
```

```
 ylim(60, 250) +
 scale_color_manual(labels=c("weekdays", "weekends"), values=c("black", "red")) +
 theme_minimal() +
 theme(axis.text.x=element_text(angle=60, hjust=1), legend.title=element_blank())
print(pickups)
```

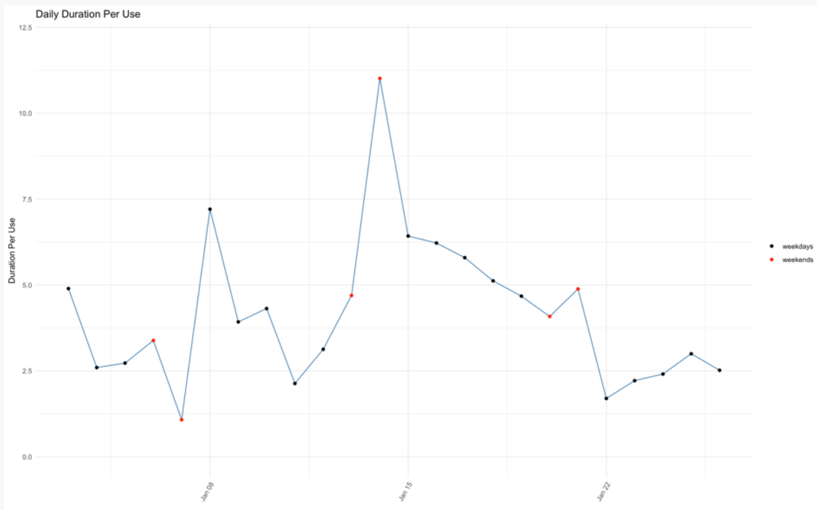



```
proportion = ggplot(screen_data, aes(x=Date, y=Proportion.Social.ST, color=if_weekend)) +
 geom_line(color="steelblue") +
 geom_point() +
 xlab("") + ylab("Proportion") +
 ggtitle("Daily Proportion of Social Screen Time")+
 ylim(0, 1) +
 scale_color_manual(labels=c("weekdays", "weekends"), values=c("black", "red")) +
 theme_minimal() +
 theme(axis.text.x=element_text(angle=60, hjust=1), legend.title=element_blank())
print(proportion)
```




```
duration = ggplot(screen_data, aes(x=Date, y=Duration.Per.Use, color=if_weekend)) +
 geom_line(color="steelblue") +
```

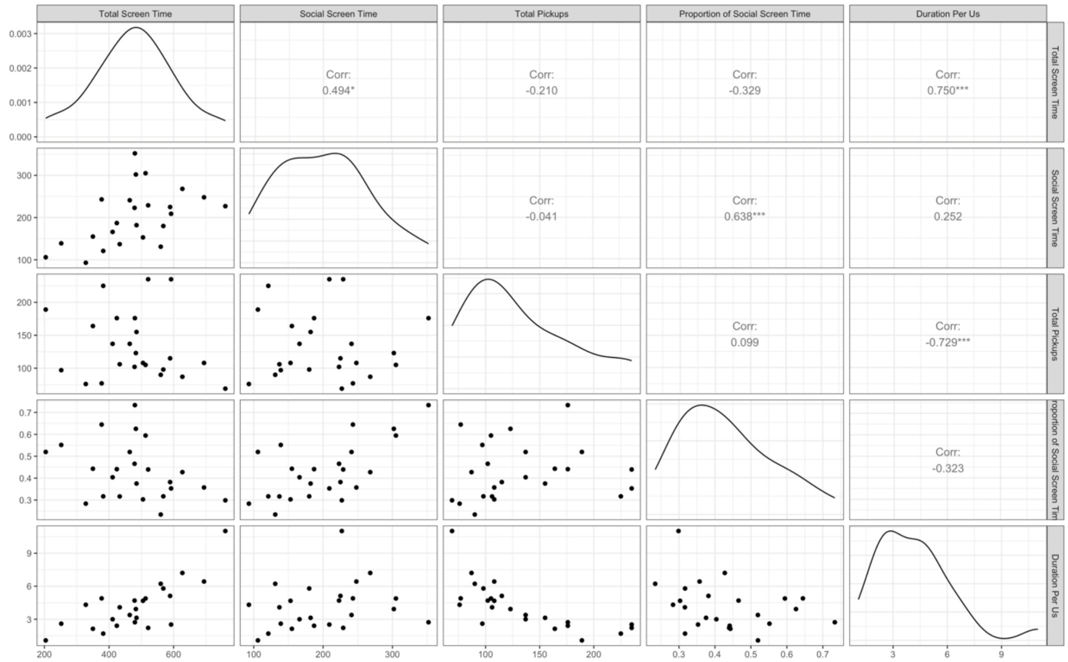
```r
 geom_point() +
 xlab("") + ylab("Duration Per Use") +
 ggtitle("Daily Duration Per Use")+
 ylim(0, 12) +
 scale_color_manual(labels=c("weekdays", "weekends"), values=c("black", "red")) +
 theme_minimal() +
 theme(axis.text.x=element_text(angle=60, hjust=1), legend.title=element_blank())
print(duration)
```



(b)

There is a significant positive correlation (corr = 0.750) between total screen time and the duration per use, which is also the highest among these pairs. Meanwhile, there is a significant negative correlation (corr = -0.729) between total pickups and the duration per use. The distribution of total screen time tends to be normal, while all other four tend to be right skewed.

```r
library(GGally)
ggpairs(screen_data, columns=c("Total.ST.min.true", "Social.ST.min.true", "Pickups", "Proportion.Social.ST", "Duration.Per.Use"),
      columnLabels=c("Total Screen Time", "Social Screen Time", "Total Pickups", "Proportion of Social Screen Time", "Duration Per Us")) +
 theme_bw()
```

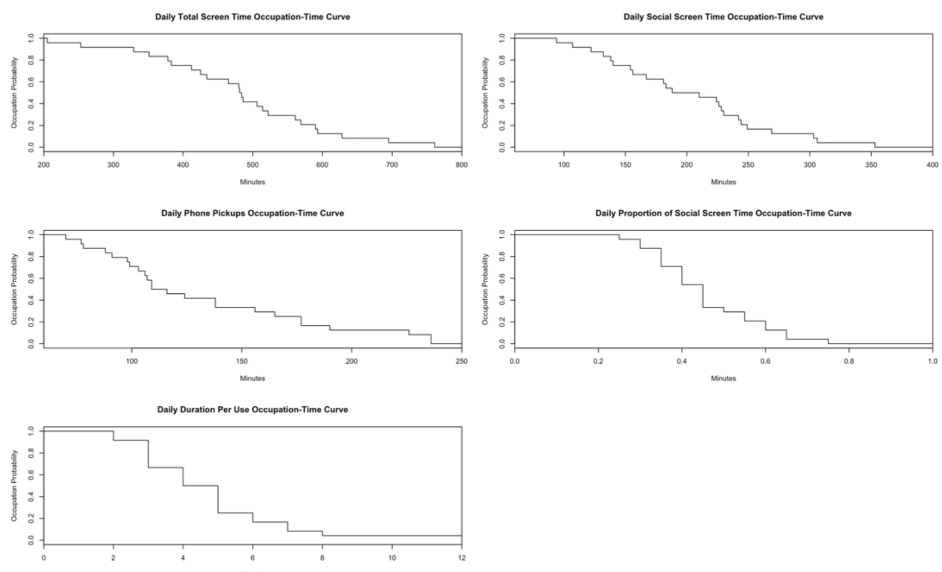


(c)

The pattern of the occupation-time curve of total screen time and social screen time are similar. Both have obvious leaps, indicating that a particular time of usage of the screen is more common than others, with a slight correlation between them. The curve of pickups has a larger leap between 60 to 100, and small leaps between 100 to 150, and so on, which indicates that the number of pickups is more commonly to be between 60 to 100 times per day. We can also tell this from the right skewed plot of total pickups. The curve of the proportion of social screen has a significant decreasing leap, which corresponds to the skewness of the distribution of the proportion of social screen. Same for the duration per use, steep leaps align with the skewness of the distribution plot.

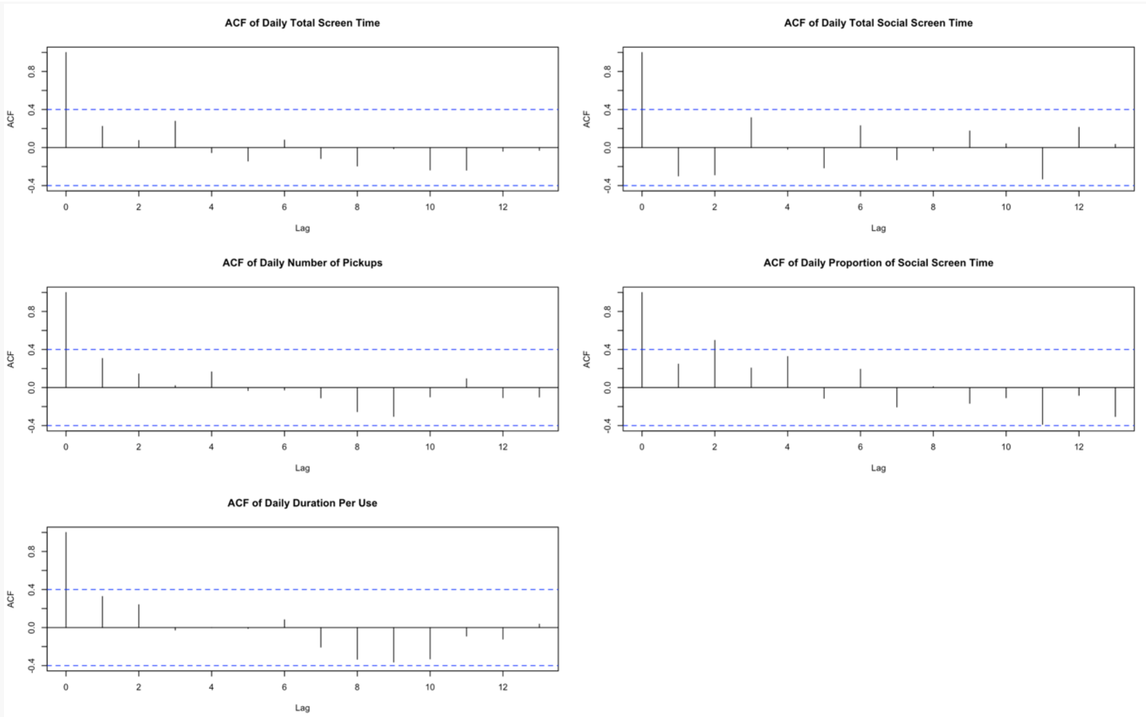
```
calculate_occupation_prob <- function(data_column, thresholds) {
  sapply(thresholds, function(x) mean(data_column >= x))
}

par(mfrow = c(3, 2))
thresholds1 = seq(200, 800, by = 1)
occupation_prob1 <- calculate_occupation_prob(screen_data$Total.ST.min.true, thresholds1)
plot(thresholds1, occupation_prob1, type = "s", xlab = "Minutes", ylab = "Occupation Probability",
     main = "Daily Total Screen Time Occupation-Time Curve", xlim = c(min(thresholds1),
max(thresholds1)), ylim = c(0, 1), xaxs = "i")

thresholds2 = seq(60, 400, by = 1)
occupation_prob2 <- calculate_occupation_prob(screen_data$Social.ST.min.true, thresholds2)
plot(thresholds2, occupation_prob2, type = "s", xlab = "Minutes", ylab = "Occupation Probability",
     main = "Daily Social Screen Time Occupation-Time Curve", xlim = c(min(thresholds2),
max(thresholds2)), ylim = c(0, 1), xaxs = "i")

thresholds3 = seq(60, 250, by = 1)
occupation_prob3 <- calculate_occupation_prob(screen_data$Pickups, thresholds3)
plot(thresholds3, occupation_prob3, type = "s", xlab = "Minutes", ylab = "Occupation Probability",
     main = "Daily Phone Pickups Occupation-Time Curve", xlim = c(min(thresholds3),
max(thresholds3)), ylim = c(0, 1), xaxs = "i")
```

```
thresholds4 = seq(0, 1, by = 0.05)
occupation_prob4 <- calculate_occupation_prob(screen_data$ Proportion.Social.ST, thresholds4)
plot(thresholds4, occupation_prob4, type = "s", xlab = "Minutes", ylab = "Occupation Probability",
    main = "Daily Proportion of Social Screen Time Occupation-Time Curve", xlim =
c(min(thresholds4), max(thresholds4)), ylim = c(0, 1), xaxs = "i")

thresholds5 = seq(0, 12, by = 1)
occupation_prob5 <- calculate_occupation_prob(screen_data$Duration.Per.Use, thresholds5)
plot(thresholds5, occupation_prob5, type = "s", xlab = "Minutes", ylab = "Occupation Probability",
    main = "Daily Duration Per Use Occupation-Time Curve", xlim = c(min(thresholds5),
max(thresholds5)), ylim = c(0, 1), xaxs = "i")
```



(d)

```
library(forecast)

par(mfrow = c(3, 2))

acf_1 = acf(screen_data$Total.ST.min.true, plot = TRUE, main="ACF of Daily Total Screen Time")
print(acf_1)

acf_2 = acf(screen_data$Social.ST.min.true, plot = TRUE, main="ACF of Daily Total Social Screen Time
")
print(acf_2)

acf_3 = acf(screen_data$Pickups, plot = TRUE, main="ACF of Daily Number of Pickups")
print(acf_3)

acf_4 = acf(screen_data$Proportion.Social.ST, plot = TRUE, main="ACF of Daily Proportion of Social S
creen Time")
print(acf_4)

acf_5 = acf(screen_data$Duration.Per.Use, plot = TRUE, main="ACF of Daily Duration Per Use")
print(acf_5)
```




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

     0      1      2      3      4      5      6      7      8      9     10     11     12     13 
 1.000  0.223  0.075  0.277 -0.054 -0.142  0.079 -0.116 -0.195 -0.013 -0.237 -0.239 -0.039 -0.030 

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

     0      1      2      3      4      5      6      7      8      9     10     11     12     13 
 1.000 -0.299 -0.288  0.314 -0.019 -0.214  0.230 -0.130 -0.033  0.175  0.039 -0.331  0.213  0.034 

Autocorrelations of series ‘screen_data$Pickups’, by lag

     0      1      2      3      4      5      6      7      8      9     10     11     12     13 
 1.000  0.307  0.144  0.020  0.165 -0.032 -0.026 -0.109 -0.255 -0.304 -0.099  0.093 -0.107 -0.100 

Autocorrelations of series ‘screen_data$Proportion.Social.ST’, by lag

     0      1      2      3      4      5      6      7      8      9     10     11     12     13 
 1.000  0.247  0.496  0.206  0.326 -0.113  0.193 -0.205  0.010 -0.166 -0.108 -0.388 -0.082 -0.305 

Autocorrelations of series ‘screen_data$Duration.Per.Use’, by lag

     0      1      2      3      4      5      6      7      8      9     10     11     12     13 
 1.000  0.327  0.241 -0.026  0.000 -0.010  0.082 -0.206 -0.334 -0.363 -0.330 -0.090 -0.122  0.036 
```

Problem 3

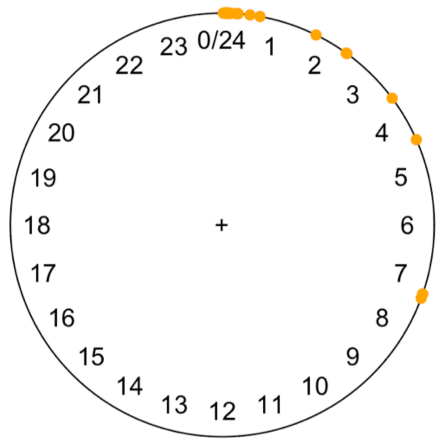
(a)

```
library(circular)
screen_data <- screen_data %>%
 mutate(Pickup.1st.angular = (hour(Pickup.1st) * 60 + minute(Pickup.1st)) / (24 * 60) * 360)
```

(b)

The most frequent time points in between 0:00AM and 1:00AM, which obviously shows my rest schedule during winter break, since I normally went to bed between 0:00AM and 0:30AM. Meanwhile, there are a few points shows to be between 2:00AM and 6:00AM, which is the time that I naturally wake due to the jet lag after I first got back here from China.

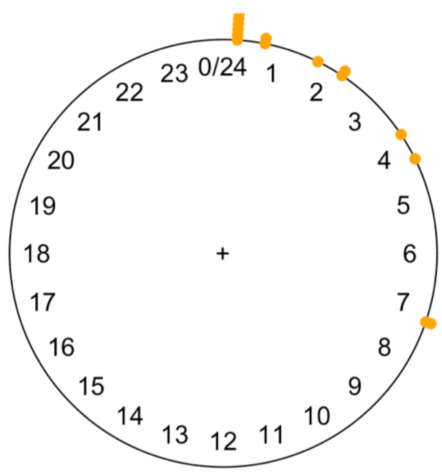
```
first.pickup.cir = circular(screen_data$Pickup.1st.angular, units="degrees", template="clock24")
plot(first.pickup.cir, col="orange")
```



(c)

I chose the size of the bin to be 48 which makes 30 minutes to be within one interval.

```
plot(first.pickup.cir, stack=TRUE, bins=48, col="orange")
```



Problem 4

(a)

The daily total screen time St is included in this Poisson model to represent the proportional relationship between the duration of screen use and the expected time of pickups. An increase in the time of using phones leads to a higher likelihood of more times of pickups.

(b)

```r
screen_data$Total.ST.hours <- screen_data$Total.ST.min.true / 60
model <- glm(Pickups ~ offset(log(Total.ST.hours)), family = poisson(link = "log"), data = screen_dat
a)
summary(model)
```

```
Call:
glm(formula = Pickups ~ offset(log(Total.ST.hours)), family = poisson(link = "log"),
    data = screen_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.81367    0.01771   158.9   <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: 800.48  on 23  degrees of freedom
Residual deviance: 800.48  on 23  degrees of freedom
AIC: 962.42

Number of Fisher Scoring iterations: 4
```

(c)

```r
screen_data$Xt <- ifelse(screen_data$if_weekend == FALSE, 1, 0)
screen_data$Zt <- ifelse(screen_data$Date >= as.Date("2024-01-10"), 1, 0)
model_c <- glm(Pickups ~ Xt + Zt + offset(log(Total.ST.hours)), family = poisson(link = "log"), data =
 screen_data)
summary(model_c)
```

```
Call:
glm(formula = Pickups ~ Xt + Zt + offset(log(Total.ST.hours)),
    family = poisson(link = "log"), data = screen_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.80014    0.04766  58.749  < 2e-16 ***
Xt           0.14459    0.04262   3.393 0.000692 ***
Zt          -0.13198    0.03954  -3.338 0.000845 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 800.48  on 23  degrees of freedom
Residual deviance: 777.21  on 21  degrees of freedom
AIC: 943.15

Number of Fisher Scoring iterations: 5
```

1. There is significant evidence at the $\alpha = 0.05$ level that the behavior of daily pickups is different between weekdays and weekends.
2. There is significant evidence at the $\alpha = 0.05$ level that the behavior of daily pickups changed after the start of the winter semester.

However, there is a special situation on me. My flight from China back to Detroit was initially Jan.14th, 2024. But I got a terrible fever the day before my original flight (although I found that this flight was also cancelled due to the snow of Michigan), so I have to change my flight to Jan.19st. Therefore, I also tried the same steps with setting Jan.19th as my changing point of winter semester.

```
screen_data$Xt <- ifelse(screen_data$if_weekend == FALSE, 1, 0)
screen_data$Zt <- ifelse(screen_data$Date >= as.Date("2024-01-19"), 1, 0)
model_c <- glm(Pickups ~ Xt + Zt + offset(log(Total.ST.hours)), family = poisson(link = "log"), data =
 screen_data)
summary(model_c)
```

```
Call:
glm(formula = Pickups ~ Xt + Zt + offset(log(Total.ST.hours)),
    family = poisson(link = "log"), data = screen_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.56127    0.04048  63.275  < 2e-16 ***
Xt           0.14891    0.04261   3.495 0.000474 ***
Zt           0.37233    0.03592  10.365  < 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: 800.48  on 23  degrees of freedom
Residual deviance: 683.47  on 21  degrees of freedom
AIC: 849.41

Number of Fisher Scoring iterations: 5
```

1. There is significant evidence at the $\alpha = 0.05$ level that the behavior of daily pickups is different between weekdays and weekends.
2. There is significant evidence at the $\alpha = 0.05$ level that the behavior of daily pickups changed after the start of the winter semester.

Problem 5

(a)

```
print(mle.vonmises(screen_data$Pickups))
```

```
Call:
mle.vonmises(x = screen_data$Pickups)

mu: 0.528  ( 1.048 )

kappa: 0.2769  ( 0.2928 )
```

```
screen_data$Pickup.1st.radians <- with(screen_data, Pickup.1st.angular * (pi / 180))
first_pickups_circular <- circular(screen_data$Pickup.1st.radians, units = "radians", template = "non
e")
estimate = mle.vonmises(first_pickups_circular)
print(estimate)
```

```
Call:
mle.vonmises(x = first_pickups_circular)

mu: 0.2958  ( 0.1106 )

kappa: 3.955  ( 1.027 )
```

(b)

The probability of the first pickup being at 8:30 AM or later is approximately 0.00116, indicating it is rarely for me to pick up my phone at 8:30 AM or even later.

```
T830am <- (8.5 / 24) * 2 * pi
cum_prob = pvonmises(T830am, estimate$mu, estimate$kappa)
prob_830am_or_later = 1 - cum_prob
print(prob_830am_or_later)

## [1] 0.001159951
```

Reference:

Beierle, Felix, et al. "Frequency and duration of daily smartphone usage in relation to personality traits." *Digital Psychology* 1.1 (2020): 20-28.