HW1-620

Bulun Te

2024-01-27

Problem 1

(a)

The collection of the data is aim at understanding individual behavior using screen time of mobile device. One kind of potential interest is the relationship between porportion of time used in socializing applications with regard to total screen time and the number of total pickups.

Regarding the question, there are some literature papers indicating that the number of pickups may be related in a indirect way to the time spent on socializing applications. For example, Sun et al. (2022) investigates the underlying connections between adolescents' mobile phone addiction and psychological factors like social anxiety and impulsivity. The findings suggest a complex interplay where a "grades-ranking-first" mentality, combined with social anxiety, drives increased mobile phone usage.Anna et al (2021) finds a significant and positive correlation between social anxiety and smartphone addiction, indicating that social anxiety might be a contributing factor to increased smartphone usage.

Based on past research, we can hypothesize that the number of pickups might be correlated with the proportion of time spent on socializing applications in an indirect way which might be able to be modelled using statistical tools.

(b)

?

(c)

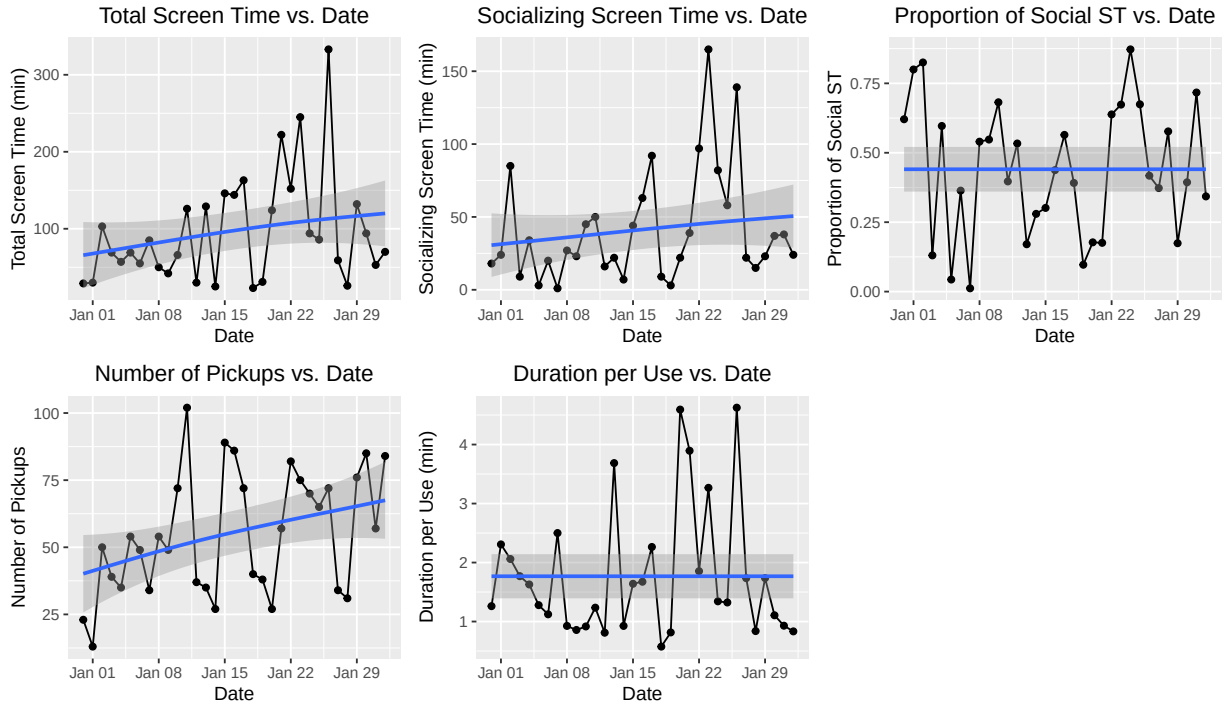
The data is collected through screen time usage monitored by personal iphone, including total screen time of a day, total time spent on socializing applications, the number of pickup times in day and the fisrt time to pickup in a day. The data is collected from 12/31/2023 to 1/26/2024 with 27 data points. The data is collected by the iphone itself and the data is collected from the iphone user.

(d)

```
## [1] "en_US.UTF-8"
```

Problem 2

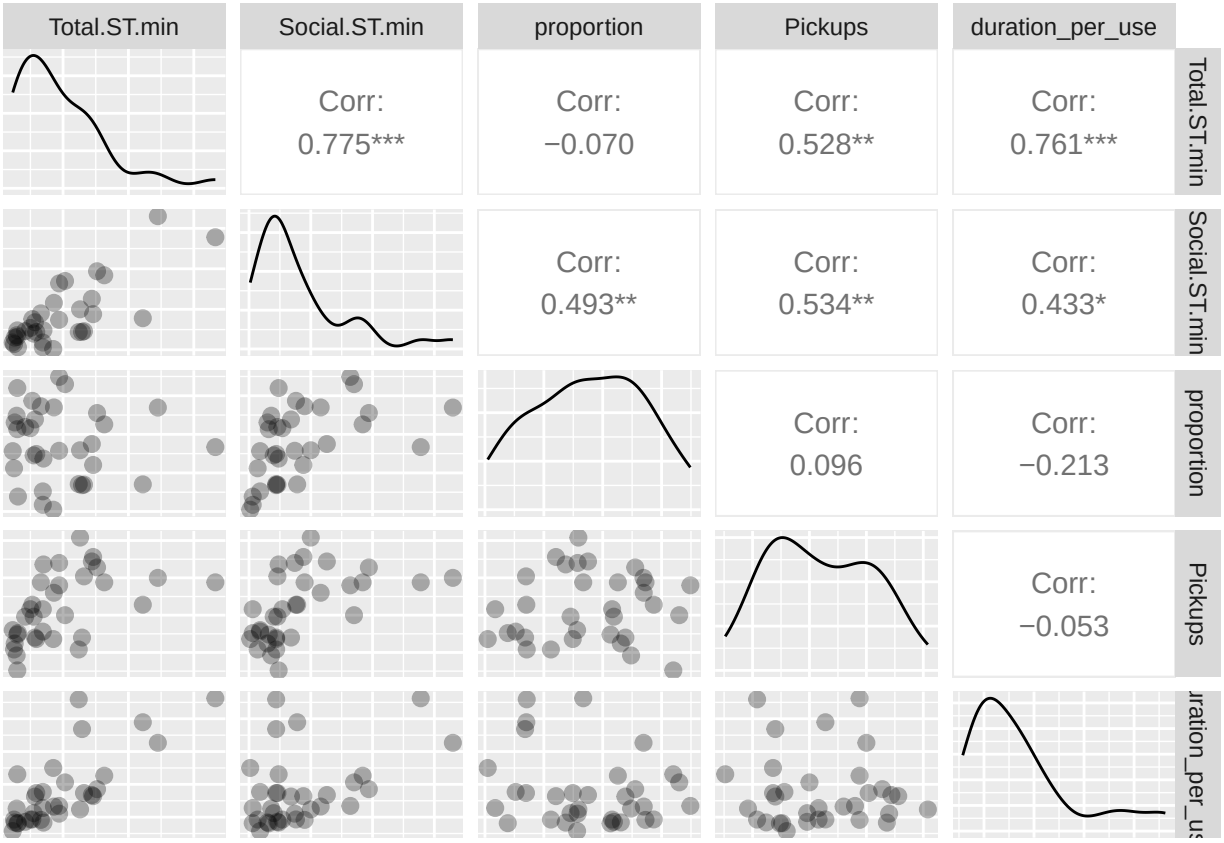
(a)



Comment:

1. From the plot we could see that, both the total screen time, pick up times, duration per use and total socializing screen time are increasing over time.
2. The Duration per Use vs. Date plot shows with some spikes that could be outliers or days of particularly long usage sessions.
3. After generalized additive smoothing, unlike the others, this metric fluctuates more without a clear trend, as indicated by the linear model trend line.

(b)



Comment:

From the scatter plot, it tells that:

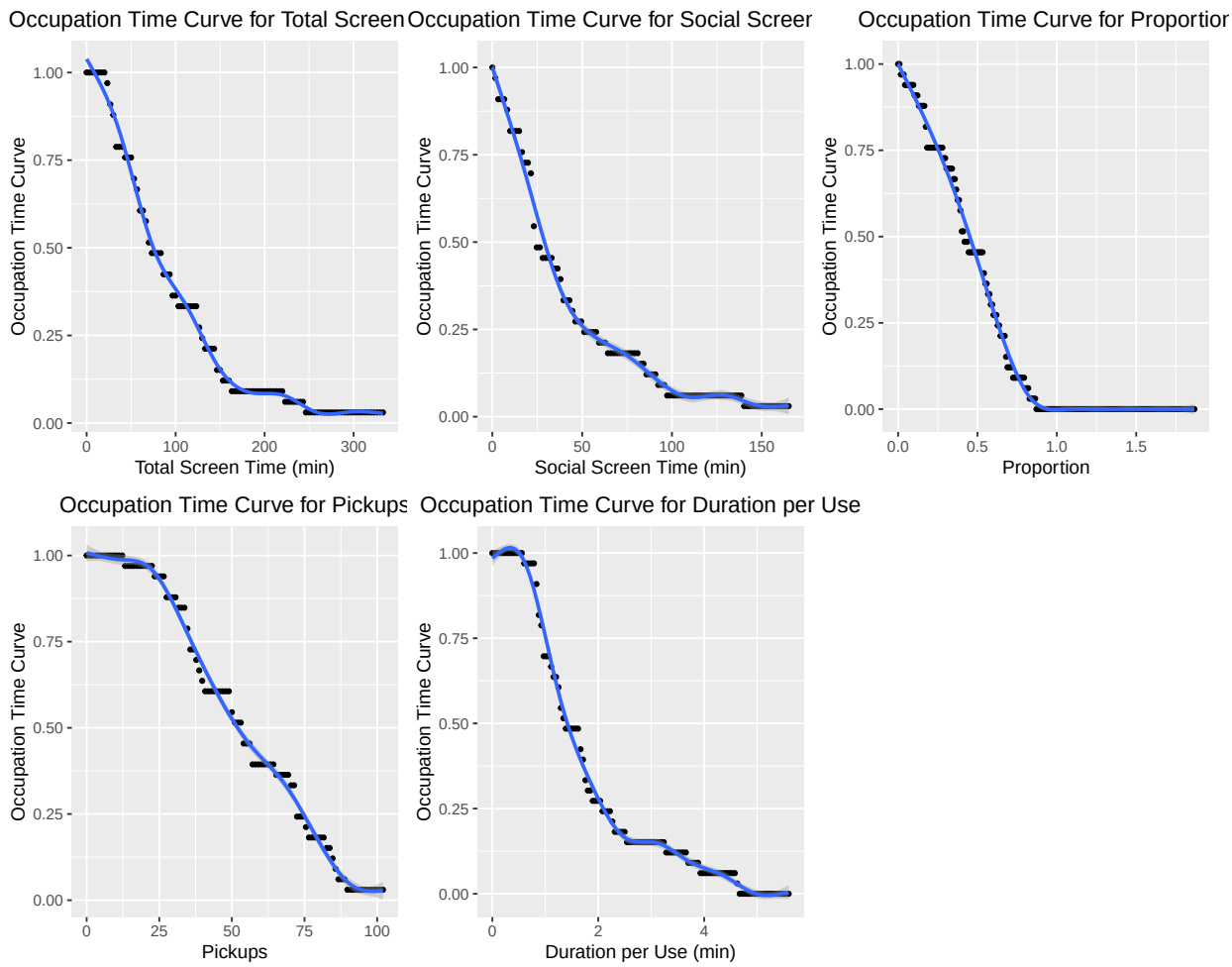
1. The total screen time and socializing screen time have highest correlation, which is reasonable since socializing screen time is a part of total screen time.

2.Social Screet time and pickups are also very correlated, indicating that most of the pickups might comes from socializing applications.

3. The distributions of Total.ST.min, Social.ST.min, proportion and Pickups are all right skewed.

(c)

Occupation Time Curve for five time variables



Comment:

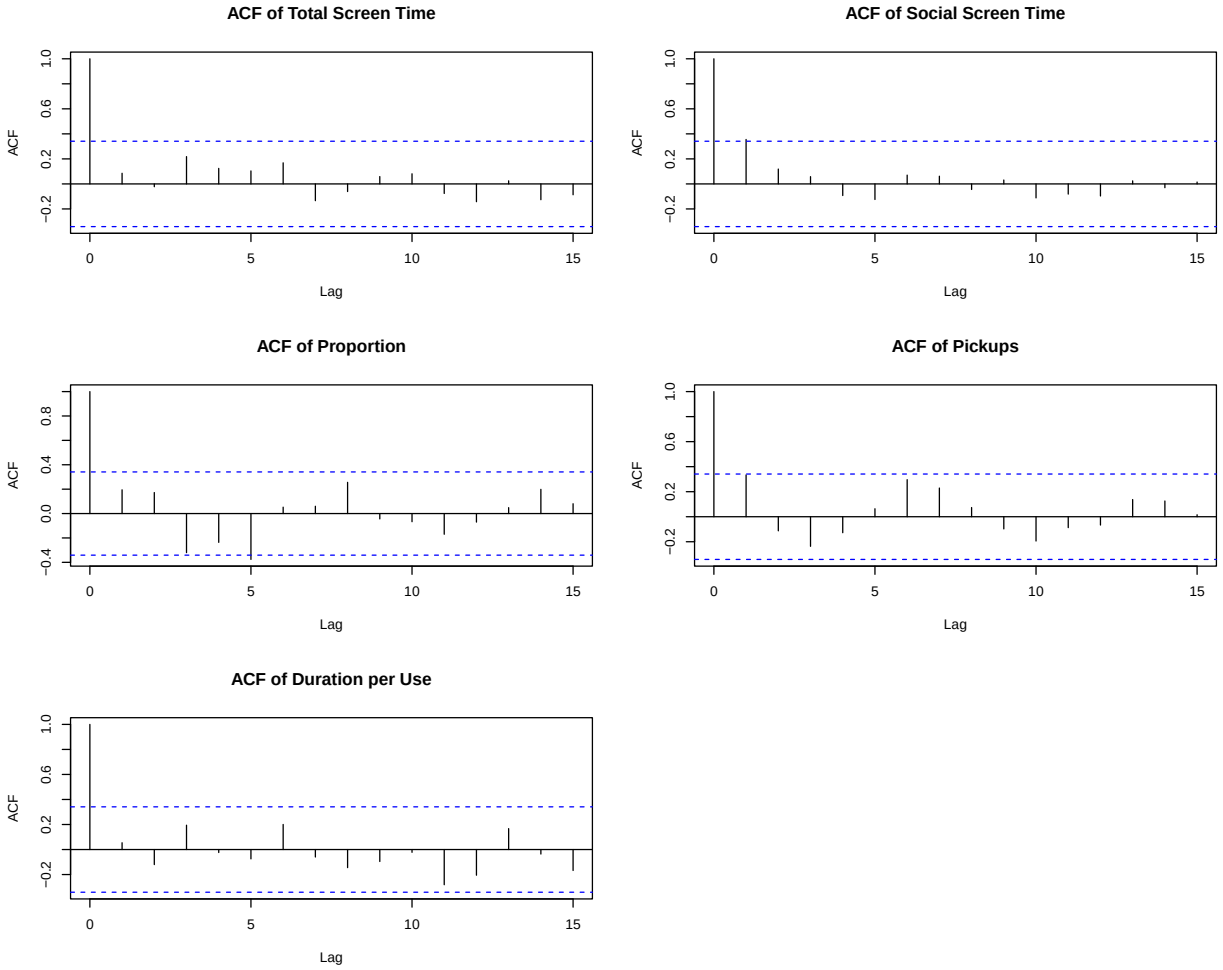
From the occupation time plots of Total Screen time, total social screen time, and pickups, the individual's pattern on screen time usage and time spent on socializing applications are relatively scattered.

The occupation time curve of proportion and duration per use are more concentrated.

Table 1: ACF values for five variables

lag	acf_total	acf_social	acf_proportion	acf_pickups	acf_duration
0	1.0000	1.0000	1.0000	1.0000	1.0000
1	0.0852	0.3554	0.1947	0.3339	0.0542
2	-0.0224	0.1188	0.1723	-0.1129	-0.1204
3	0.2183	0.0581	-0.3205	-0.2366	0.1945
4	0.1240	-0.0929	-0.2364	-0.1279	-0.0240
5	0.1039	-0.1245	-0.3761	0.0638	-0.0749
6	0.1685	0.0703	0.0524	0.2963	0.2004
7	-0.1340	0.0618	0.0600	0.2293	-0.0599
8	-0.0610	-0.0445	0.2560	0.0735	-0.1449
9	0.0588	0.0312	-0.0441	-0.0975	-0.0958

(d)

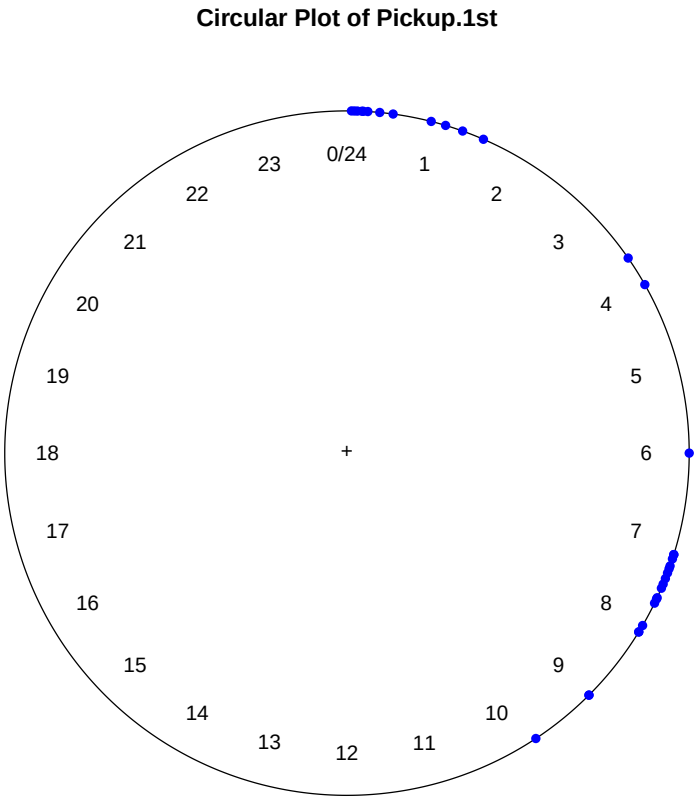


Comment:

From the autocorrelation plots, only significant spike displayed at lag 0 and all other lags are within the 95% confidence interval for all five variables. This indicates that there is no significant auto correlations at any lag for those series at siginificancy level of 5%.

Problem 3

(a) (b)

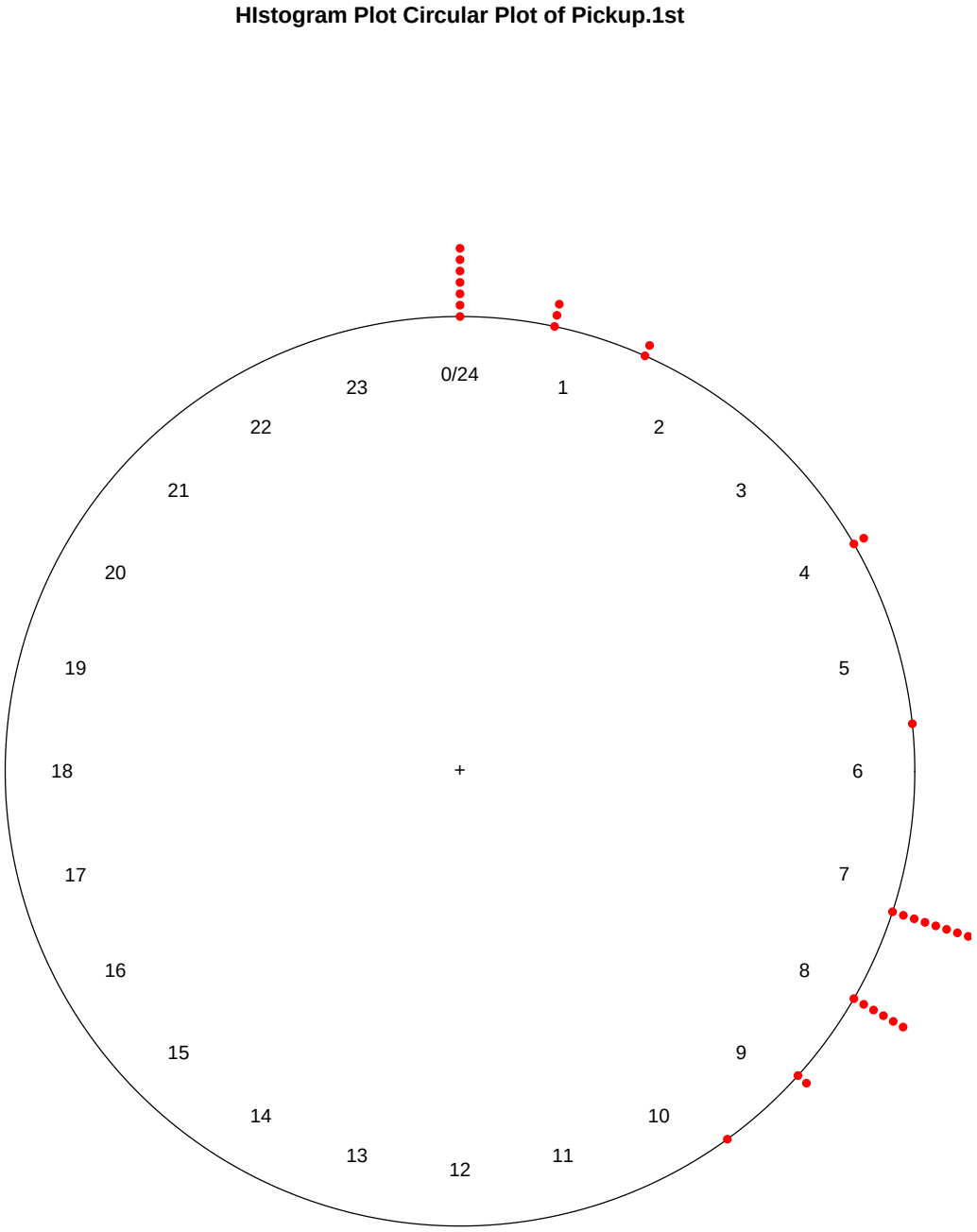



```
## $zero
## [1] 1.570796
##
## $rotation
## [1] "clock"
##
## $next.points
## [1] 0.025
```

Comment

From the clock, the plot shows that most of the first pickup time is concentrated in 7am-8am where at some days, first pickup time of the mobile phone might falls in 12am-1am. This indicates that the individual might be a morning person and he/she might have a regular schedule of waking up at 7am-8am and sometimes staying up late at night.

(c)


HIstogram Plot Circular Plot of Pickup.1st
0/24
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
+

Comment

The number of bins is set to 30, which correspond to 30 minutes per bin. Typically, people divide their daily routines into 30-minute increments.

Problem 4

(a)

the total screen time $S(t)$ acts as a scaling factor in the Poisson distribution: $\lambda \times S(t)$. This product, $\lambda \times S(t)$, represents the adjusted expected number of pickups for each day, accounting for the varying time window of opportunity.

This adjustment is necessary because the probability of a pickup is assumed to be independent and constant throughout the screen time period. With varying screen times, simply using the fixed rate λ wouldn't accurately capture the daily differences in pickup opportunities. The factor $S(t)$ incorporates this variability, resulting in a time-sensitive expected number of pickups that reflects the reality of fluctuating screen time.

(b)

```
possion_model_offset = glm(Pickups~offset(log(Total.ST.min)),data=data,family=poisson(link="log"))

summary(possion_model_offset)
```

```
##
## Call:
## glm(formula = Pickups ~ offset(log(Total.ST.min)), family = poisson(link = "log"),
##     data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.55622    0.02349  -23.68   <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: 516.99  on 32  degrees of freedom
## Residual deviance: 516.99  on 32  degrees of freedom
## AIC: 708.77
##
## Number of Fisher Scoring iterations: 4
```

(c)

```
glm_loglink_offset = glm(Pickups~ weekday+semester+offset(log(Total.ST.min)),data=data,family=poisson(li

summary(glm_loglink_offset)
```

```
##
## Call:
## glm(formula = Pickups ~ weekday + semester + offset(log(Total.ST.min)),
##     family = poisson(link = "log"), data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.68991    0.06986  -9.875  < 2e-16 ***
## weekday      0.40294    0.06191   6.508 7.61e-11 ***
## semester    -0.23399    0.05671  -4.126 3.69e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 516.99  on 32  degrees of freedom
## Residual deviance: 457.40  on 30  degrees of freedom
## AIC: 653.17
##
## Number of Fisher Scoring iterations: 5
```

(c.1)

From the model, it tells that, at the significance level of 5%, the coefficient of weekday is significuanlty ($p < 0.0001$) different from 0. This indicates that the number of pickups is significantly different between weekdays and weekends. Adjusting for the total screen time and the semester, the expected number of pickups is $e^{0.40294} = 1.4962$ times higher on weekdays than on weekends, which is understandable because more emails and messages are received during weekdays than weekends.

(c.2)

From the model, it tells that, at the significance level of 5%, the coefficient of semester is significuanlty ($p < 0.0001$) different from 0. This indicates that the number of pickups is significantly different between the semesters and vacation. Adjusting for the total screen time and the weekday, the expected number of pickups is $e^{-0.23399} = 0.7914$ times lower in the semester than in the winter vacation.

Problem 5

(a)

```r
mle_estimate = mle.vonmises(data$Pickup.1st.angular, bias = TRUE)
```

```
## Warning in as.circular(x): an object is coerced to the class 'circular' using default value for the
##   type: 'angles'
##   units: 'radians'
##   template: 'none'
##   modulo: 'asis'
##   zero: 0
##   rotation: 'counter'
## conversion.circularxradians0counter2pi
```

```
mle_estimate
```

```
##
## Call:
## mle.vonmises(x = data$Pickup.1st.angular, bias = TRUE)
##
## mu: 2.095  ( 0.4432 )
##
## kappa: 0.5663  ( 0.261 )
##
## Bias correction (Best and Fisher, 1981) applied to kappa
```

(b)

```
mu = mle_estimate$mu

kappa = mle_estimate$kappa

# The proabbility that first pickup time is later than 8:30 am
# calculating with pvonmises function as follow:

1-pvonmises(8.5, mu, kappa)
```

```
## [1] 0.46854
```