How Social Apps Influence Wake-Up Times in College Life, A Federated Learning Study

Lufeiya Liu, Anran Yao

2024-02-23

Read the data

```
timedata<-read.csv("D:/project_620/Data/groupdata_01W04 - timeseries.csv")
baseline<-read.csv("D:/project_620/Data/groupdata_01W04 - baseline.csv")
baseline<-baseline[, -ncol(baseline)]
```

```
# change the format of the 1st pickup to time format, transform to angular
timedata <- timedata %>%
  mutate(Pickup.1st = strptime(Pickup.1st, format = "%H:%M"),
         Pickup.1st.angular =
           (hour(Pickup.1st)*60+minute(Pickup.1st))/(24*60)*360)

# combine the data with baseline
combined <- timedata %>% left_join(baseline, by = c("ID"="ID"))
```

data validation

```
hm_to_min = function(hm){
  unlist(lapply(hm,function(x){
    if(grepl("h", x) && !grepl("m", x)) {
      x = gsub("h", "*60", x)
      return(eval(parse(text = x)))
    }else if(grepl("h", x)){
      x = gsub("h","*60+",x)
      x = gsub("m","",x)
      return(eval(parse(text=x)));
    }else{
      return(as.numeric(gsub("m","",x)))
    }})) #convert to total minutes
}
```

```
validation = timedata %>%
  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)
combined_validated <- validation %>% left_join(baseline, by = c("ID"="ID"))
# write.csv(combined_validated,"combined_validated.csv")
```

descriptive statistics

```
##  [1] "ID"             "workmate"       "academic"       "non.academic"
##  [5] "pets"           "sex"            "age"            "course_hours"
##  [9] "degree"         "job"            "siblings"       "apps"
## [13] "devices"        "procrastination"
```

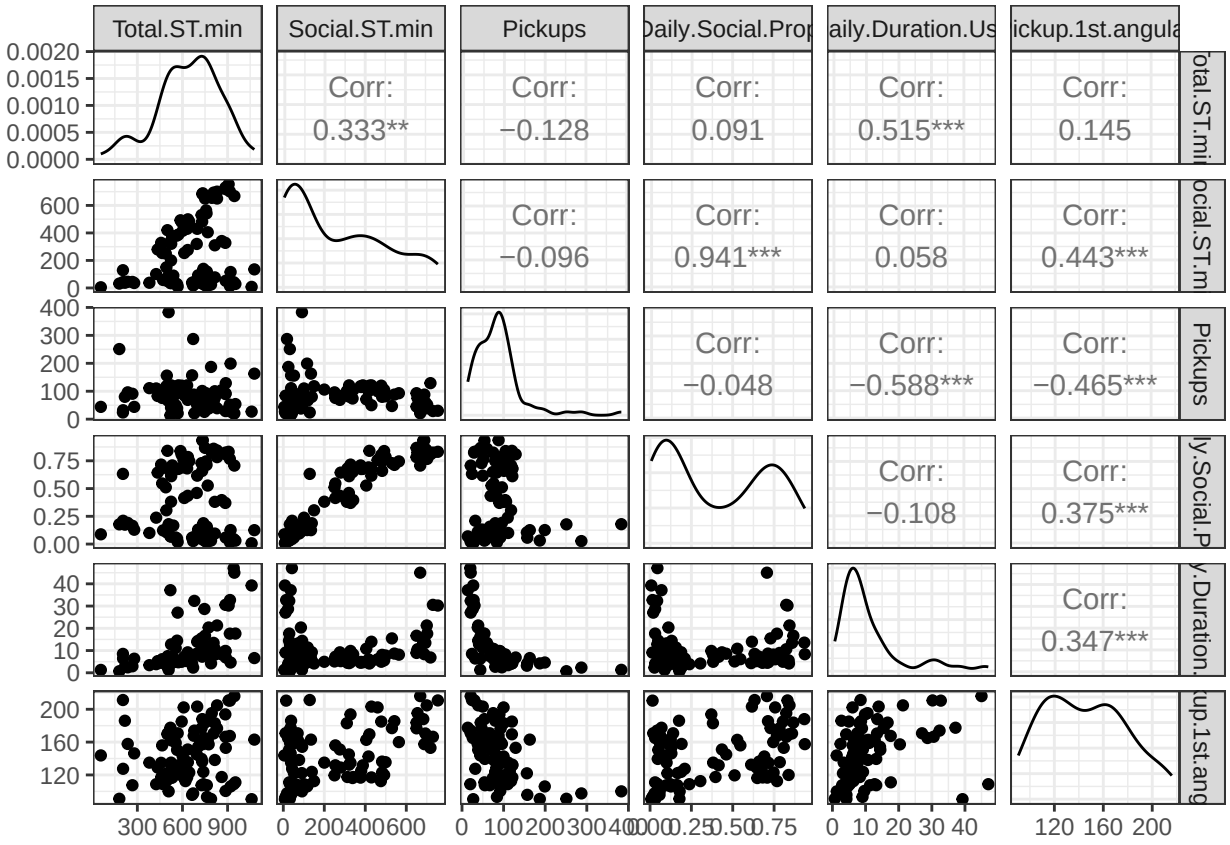
```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**Summary Statistics**
___total___	N = 2
___workmate___	0.0 (0.0)
___academic___	0.5 (0.7)
___non.academic___	1.0 (0.0)
___pets___	0.0 (0.0)
___sex___	
0	2 (100%)
___age___	22.5 (0.7)
___course_hours___	9.8 (2.5)
___degree___	
0	1 (50%)
1	1 (50%)
___job___	
1	2 (100%)
___siblings___	0.5 (0.7)
___apps___	4.0 (1.4)
___devices___	4.0 (0.0)
___procrastination___	44.5 (7.8)
___total___	N = 88
___Total.ST.min___	642.8 (207.0)
___Social.ST.min___	262.8 (238.1)
___Pickups___	87.0 (57.1)
___Pickup.1st.angular___	145.5 (33.5)
___Daily.Social.Prop___	0.4 (0.3)
___Daily.Duration.Use___	11.1 (9.8)
___Weekdays___	
0	24 (27%)
1	64 (73%)

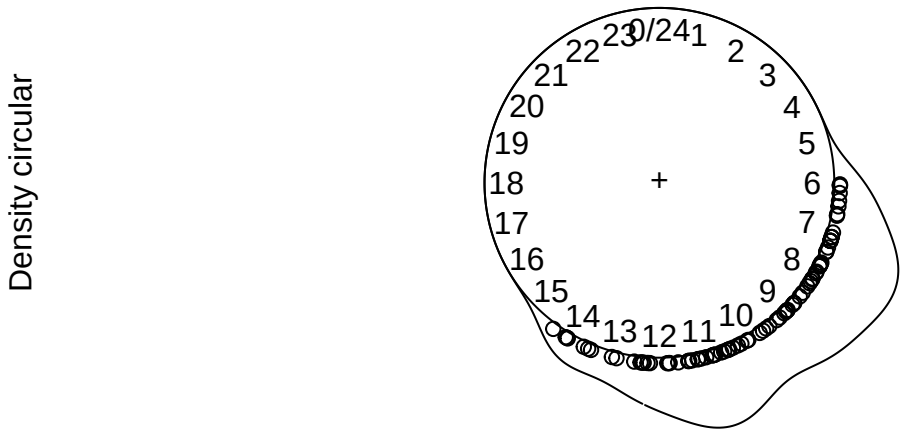
plots

correlation

```
timedata %>%
  ggpairs(columns = c(4,6,7,9,10,14), progress = FALSE)+ theme_bw()
```

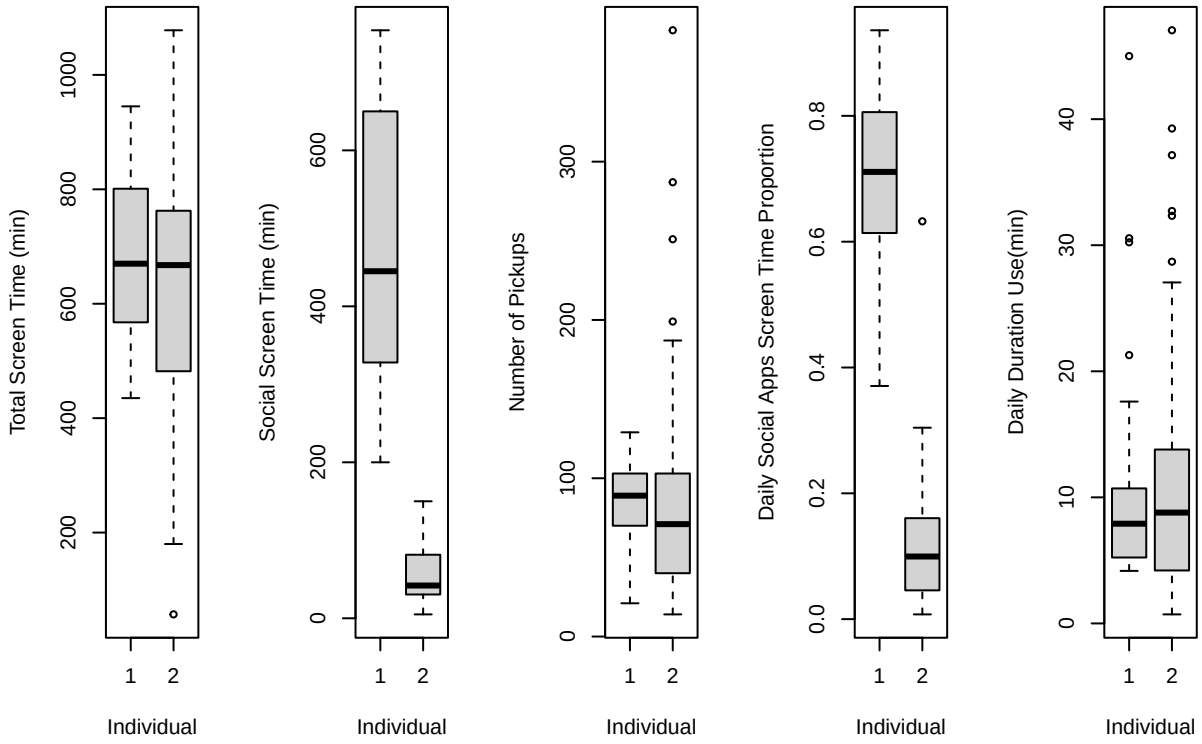



```
# first.pickup.circular distribution
layout(1)
first.pickup.cir = circular(timedata$Pickup.1st.angular, units="degrees",
                            template="clock24")
first.pickup.cir.den = density(first.pickup.cir,bw=50)
plot(first.pickup.cir.den, points.plot = T,shrink=1.3, main = "")
```



boxplots

```
layout(matrix(c(1,2,3,4,5), nrow = 1))
boxplot(Total.ST.min~ID,timedata,ylab="Total Screen Time (min)",
        xlab="Individual",all.outliers=TRUE,col = "lightgray")
boxplot(Social.ST.min~ID,timedata,ylab="Social Screen Time (min)",
        xlab="Individual",all.outliers=TRUE,col = "lightgray")
boxplot(Pickups~ID,timedata,ylab="Number of Pickups",
        xlab="Individual",all.outliers=TRUE,col = "lightgray")
boxplot(Daily.Social.Prop~ID,timedata,
        ylab="Daily Social Apps Screen Time Proportion",
        xlab="Individual",all.outliers=TRUE,col = "lightgray")
boxplot(Daily.Duration.Use~ID,timedata,ylab="Daily Duration Use(min)",
        xlab="Individual",all.outliers=TRUE,col = "lightgray")
```



federated learning

```
combined1 <- combined_validated %>%
  select(-Total.ST,-Total.ST.min,-Total.ST.match,
         -Social.ST,-Social.ST.min,-Social.ST.match) %>%
  mutate(Total.ST.min=Total.ST.min.true,
         Social.ST.min=Social.ST.min.true)
```

```
# data for person 1
timedata1 <- subset(combined1, ID==1)
# data for person 2
timedata2 <- subset(combined1, ID==2)

x1 = model.matrix(timedata1$Social.ST.min~timedata1$Pickup.1st.angular
                  +timedata1$Pickups
                  +timedata1$Weekdays
                  +timedata1$apps)
x2 = model.matrix(timedata2$Social.ST.min~timedata2$Pickup.1st.angular
                  +timedata2$Pickups
                  +timedata2$Weekdays
                  +timedata2$apps)
SSX1 = t(x1)%*%x1
SSXY1 = t(x1)%*%timedata1$Social.ST.min
SSY1 = t(timedata1$Social.ST.min) %*% timedata1$Social.ST.min
```

```
SSX2 = t(x2)%*%x2
SSXY2 = t(x2)%*%timedata2$Social.ST.min
SSY2 = t(timedata2$Social.ST.min) %*% timedata2$Social.ST.min
# beta hat estimates
beta_y <- solve(SSX1+SSX2)%*%(SSXY1+SSXY2)

sigma_y <-((SSY1+SSY2)-2*t(beta_y)%*%(SSXY1+SSXY2)+
             t(beta_y)%*%(SSX1+SSX2)%*%(beta_y))/(88-5)
ste <- sqrt(sigma_y)
# standard errors
se_beta = t(ste%*%sqrt(diag(as.matrix(solve(as.matrix(SSX1+SSX2))))))
# t statistics
t_statistic = beta_y/se_beta

# p values
p_value = c(2*(1-pt(q=abs(t_statistic),df=83)))

# federal results
fed_table = data.frame(beta_y=beta_y, se_beta=se_beta, t_statistic, p_value)
fed_table
```

```
##                                  beta_y     se_beta t_statistic      p_value
## (Intercept)                  926.6194372 106.5476425   8.6967615 2.642331e-13
## timedata1$Pickup.1st.angular   1.1139260   0.4249347   2.6214050 1.041383e-02
## timedata1$Pickups              0.1334644   0.2265048   0.5892343 5.573046e-01
## timedata1$Weekdays           -75.4676364  26.8814888  -2.8074203 6.222763e-03
## timedata1$apps              -195.5941595  12.0432888 -16.2409258 0.000000e+00
```

```
# calculate RSS
RSS = ((SSY1+SSY2)-2*t(beta_y)%*%(SSXY1+SSXY2)+
        t(beta_y)%*%(SSX1+SSX2)%*%(beta_y))
RSS
```

```
##          [,1]
## [1,] 930131.6
```

```
# calculate R^2a
y_bar1 = mean(timedata1$Social.ST.min)
y_bar2 = mean(timedata2$Social.ST.min)
y_bar = 44/88*y_bar1+44/88*y_bar2
TSS = (SSY1+SSY2)-88*y_bar^2
TSS
```

```
##         [,1]
## [1,] 4949100
```

```
Ra_2 = 1-RSS/(88-5)/(TSS/(88-1))
Ra_2 # 0.8030031
```

```
##           [,1]
## [1,] 0.8030031
```

confirmation analysis

```
lm_1 <- lm(Social.ST.min~Pickup.1st.angular
           +Pickups
           +Weekdays
           +apps , data=combined1)
(lm_1.sum<-summary(lm_1))
```

```
##
## Call:
## lm(formula = Social.ST.min ~ Pickup.1st.angular + Pickups + Weekdays +
##     apps, data = combined1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -239.672  -66.083   -7.213   70.406  250.867
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         926.6194   106.5476   8.697 2.64e-13 ***
## Pickup.1st.angular    1.1139     0.4249   2.621  0.01041 *
## Pickups               0.1335     0.2265   0.589  0.55730
## Weekdays            -75.4676    26.8815  -2.807  0.00622 **
## apps               -195.5942    12.0433 -16.241  < 2e-16 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 105.9 on 83 degrees of freedom
## Multiple R-squared:  0.8121, Adjusted R-squared:  0.803
## F-statistic: 89.66 on 4 and 83 DF,  p-value: < 2.2e-16
```

```
anova(lm_1) # see RSS
```

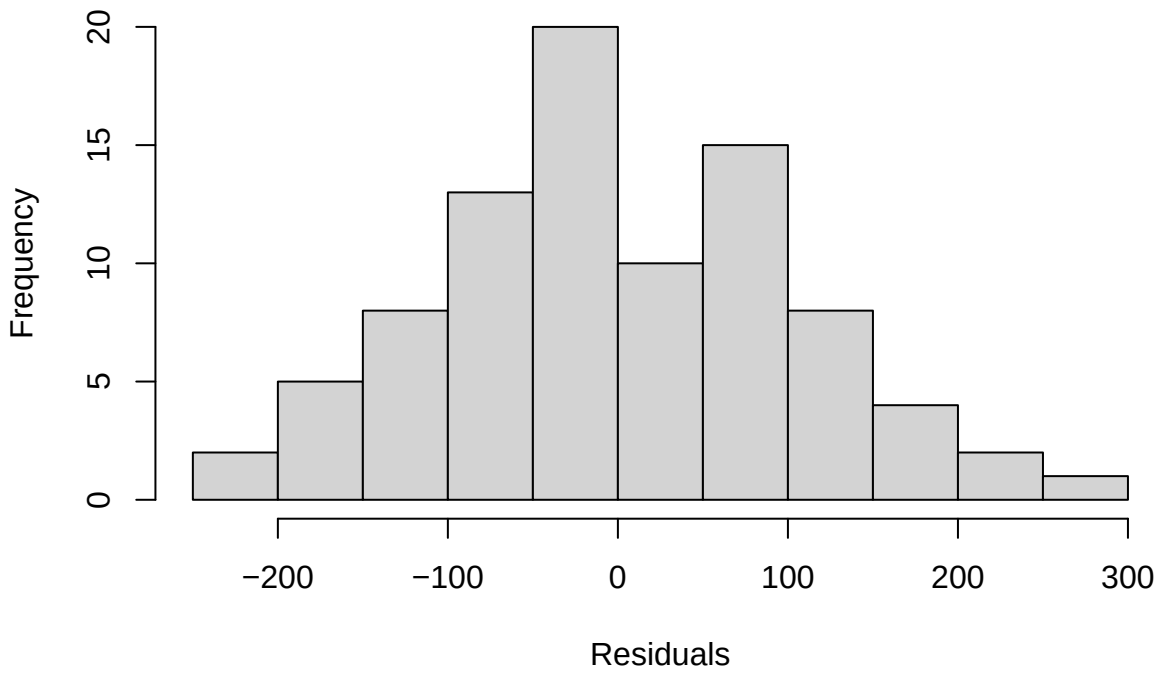
```
## Analysis of Variance Table
##
## Response: Social.ST.min
##                    Df  Sum Sq Mean Sq  F value    Pr(>F)
## Pickup.1st.angular  1  975353  975353  87.0353 1.433e-14 ***
## Pickups             1   75139   75139   6.7050   0.01135 *
## Weekdays            1   12589   12589   1.1234   0.29227
## apps                1 2955887 2955887 263.7677 < 2.2e-16 ***
## Residuals          83  930132   11206
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
lm_1.sum$coefficients
```

```
##                      Estimate  Std. Error     t value     Pr(>|t|)
## (Intercept)         926.6194372 106.5476425   8.6967615 2.641351e-13
## Pickup.1st.angular    1.1139260   0.4249347   2.6214050 1.041383e-02
## Pickups               0.1334644   0.2265048   0.5892343 5.573046e-01
## Weekdays            -75.4676364  26.8814888  -2.8074203 6.222763e-03
## apps               -195.5941595  12.0432888 -16.2409258 1.694520e-27
```

model diagnosis

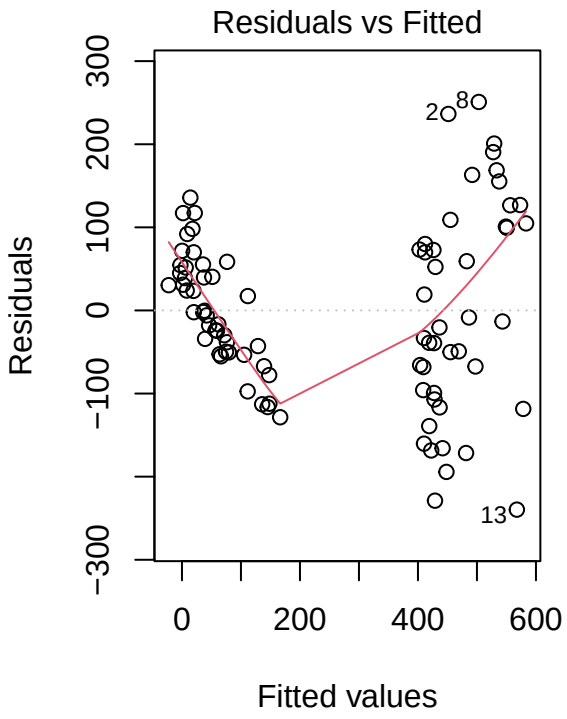
```
hist(lm_1$residuals,main = "",xlab = "Residuals")
```

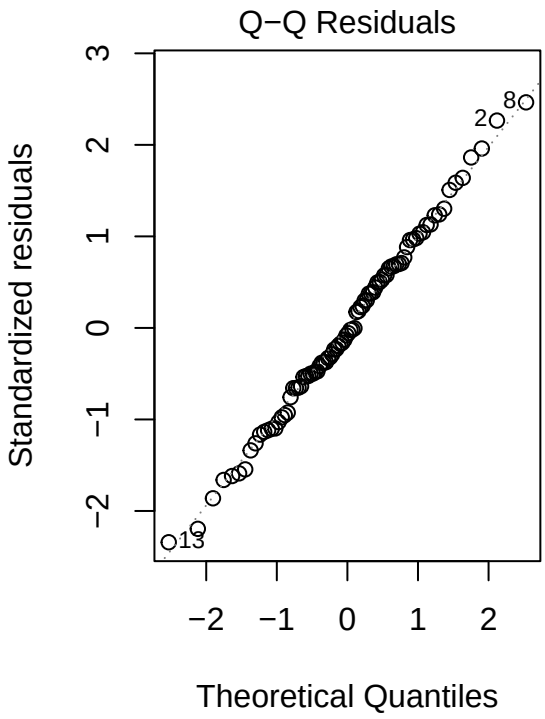



```
vif(lm_1)%>% kbl(format = "latex",booktabs = TRUE,longtable=TRUE,digits = 2)
```

	x
Pickup.1st.angular	1.57
Pickups	1.30
Weekdays	1.13
apps	1.14

```
par(mfrow = c(1, 2))
plot(lm_1,c(1,2))
```


Residuals vs Fitted
Residuals
Fitted values
−300
−100
0
100
200
300
0
200
400
600
2
8
13


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−2
−1
0
1
2
3
2
8
13