

TIME I SPEND ON COURSE-WORK AND SOCIAL-MEDIA

Lorraine Oloo

Due Friday, March 19 by 5:00 PM EST

How do I spend my time?

Data wrangling

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.0.6      v dplyr  1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

library(ical)
library(readxl)
library(vctrs)

##
## Attaching package: 'vctrs'

## The following object is masked from 'package:dplyr':
##
##   data_frame
```

```
## The following object is masked from 'package:tibble':  
##  
## data_frame
```

```
library(tidyr)
```

```
#calling the data set and setting start time an end time for my times
```

```
my_calendar0 <-  
ical_parse_df("C:/Users/Lorraine/Dropbox/Data Science/Data-Science/Calender/loloo23@amherst.edu.ical/Da
```

```
  mutate(start_datetime = with_tz(start, tzone = "America/New_York")  
    , end_datetime = with_tz(end, tzone = "America/New_York")  
    , length_hour = end_datetime - start_datetime  
    , date = floor_date(start_datetime, unit = "day"))
```

```
#arrange data in order of time
```

```
my_calendar1 <- arrange(my_calendar0, date)
```

```
#combining the variables in summary
```

```
my_calendar1 <- mutate(  
  my_calendar1, summary_correct = str_trim(summary))
```

```
#creating day of the week
```

```
my_calendar1 <- mutate(my_calendar1, day_of_week = weekdays(date))
```

```
#Distinguishing between assignment and reading and the specific courses
```

```
#Study type
```

```
my_calendar1 <- my_calendar1 %>%  
  mutate(study = case_when(str_detect(summary_correct, "R") ~ "Reading",  
    str_detect(summary_correct, "A") ~ "Assignment",  
    TRUE ~ summary_correct))
```

```
#Renaming Courses
```

```
my_calendar1 <- my_calendar1 %>%  
  mutate(course = case_when(str_detect(summary_correct, "D") ~ "Data Science",  
    str_detect(summary_correct, "E") ~ "Econometrics",  
    str_detect(summary_correct, "M") ~ "Money&Banking",  
    TRUE ~ summary_correct))
```

```
#Weekday or weekend
```

```
my_calendar1 <- my_calendar1 %>%  
  mutate(type_day = ifelse(day_of_week %in% c("Saturday", "Sunday"),  
    "Weekend", "Weekday"))
```

```
#selecting variables I'll be working with
```

```
my_calendar2 <- my_calendar1 %>%  
  select(summary_correct, date, day_of_week, type_day, course, study, length_hour)
```

```
social_media <- read_excel("Social Media.xlsx")
```

```
social_media1 <- arrange(social_media, date)
```

```
#remove PU from the data set
```

```

social_media1 <- filter(social_media1, summary_correct != "PU")
#Adding day of the week
social_media1 <- mutate(social_media1, day_of_week = weekdays(date))
#Weekday or weekend
social_media1 <- social_media1 %>%
  mutate(type_day = ifelse(day_of_week %in% c("Saturday", "Sunday"),
                           "Weekend", "Weekday"))

#combine data
my_calendar3 <- my_calendar2 %>%
  mutate(length_hour = as.numeric(length_hour))

social_media2 <- social_media1 %>%
  mutate(length_hour= as.numeric(length_hour))%>%
  select(summary_correct, length_hour, day_of_week, type_day)

my_calendar3 <- select(my_calendar3,
                      c(summary_correct, length_hour, day_of_week, type_day))

total <- bind_rows(social_media2, my_calendar3)

```

Data collection process, question of interests, and variables

On a typical day, other than attending my classes, I spend most of my time doing assignments, reading and perusing through social media.

I was interested in:

1. knowing how much time, on average, I spend doing my assignments and readings per course and the two social media apps I frequently use.
2. Difference in time spent on these activities on the weekdays and weekends.

I collected the data on time spent on my courses via Google calender after the activity was done.

I collected data on time spent on my frequently used apps by manually entering the minutes spent on this activities into an Excel spread-sheet from my phone usage database.

The time frame was between 28/02/2021 and 15/03/2021

The variables in my data-base are:

A-DS - Doing Data-Science assignment
A-EC - Doing Econometrics assignment
A-MB - Doing Money and Banking assignment
IG - Instagram
WA - WhatsApp

Graphs for my studies

Data visualization 1

The bar graph shows how much time in minutes I spend on average on each subject per day over the course of the 15 days.

```
#Table for mean minutes per course per day
calender3 <- my_calendar2%>%
  group_by(day_of_week,course)%>%
  summarise(avg_min = mean(length_hour))
```

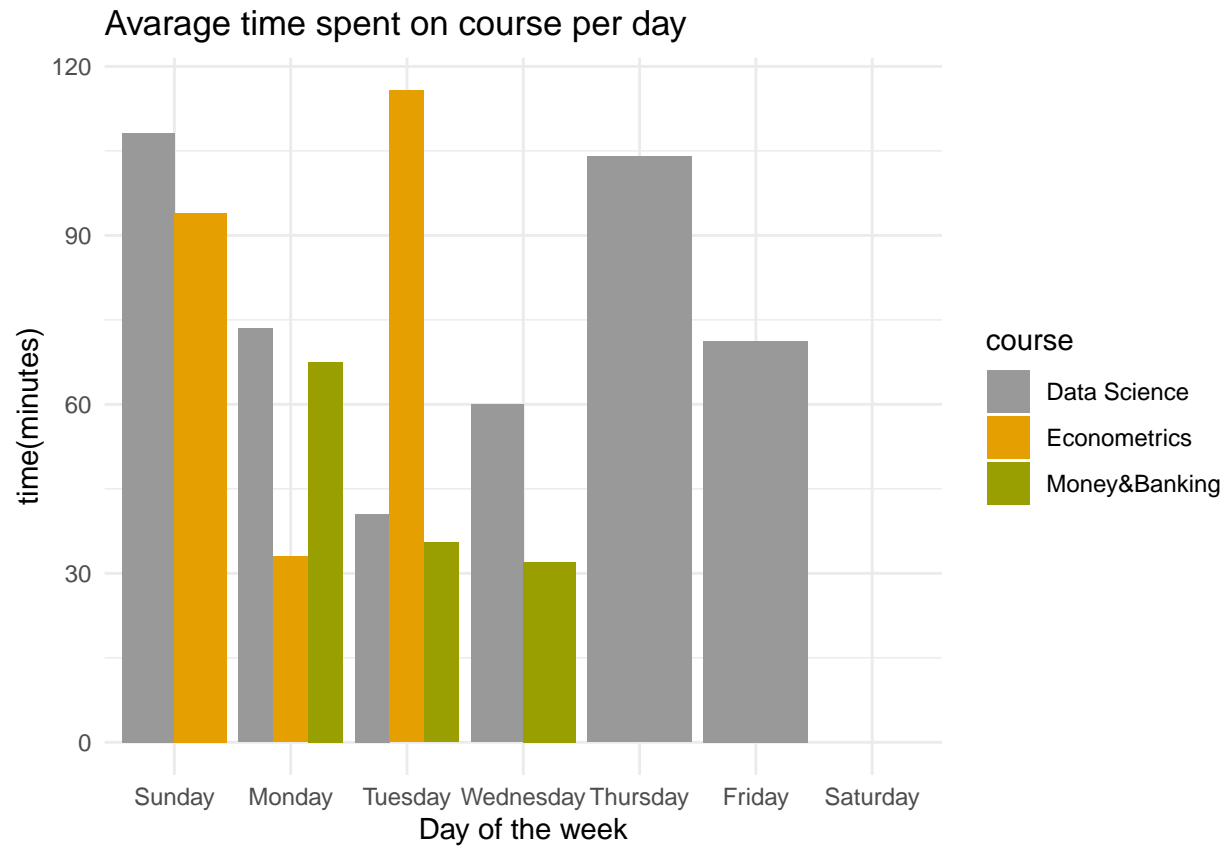
'summarise()' has grouped output by 'day_of_week'. You can override using the '.groups' argument.

```
calender3
```

```
#Bargraph showing the different time per subject per day
p<-ggplot(data=calender3,
  aes(x=day_of_week, y=avg_min, fill = course)) +
  geom_bar(stat="identity", position=position_dodge())+
  theme_minimal() +
  labs(title = "Avarage time spent on course per day",
  y="time(minutes)",
  x="Day of the week")

p + scale_fill_manual(values=c('#999999','#E69F00','#999F00'))+ scale_x_discrete(limits=c("Sunday", "Monday",
  "Tuesday", "Wednesday",
  "Thursday", "Friday",
  "Saturday"))
```

Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.



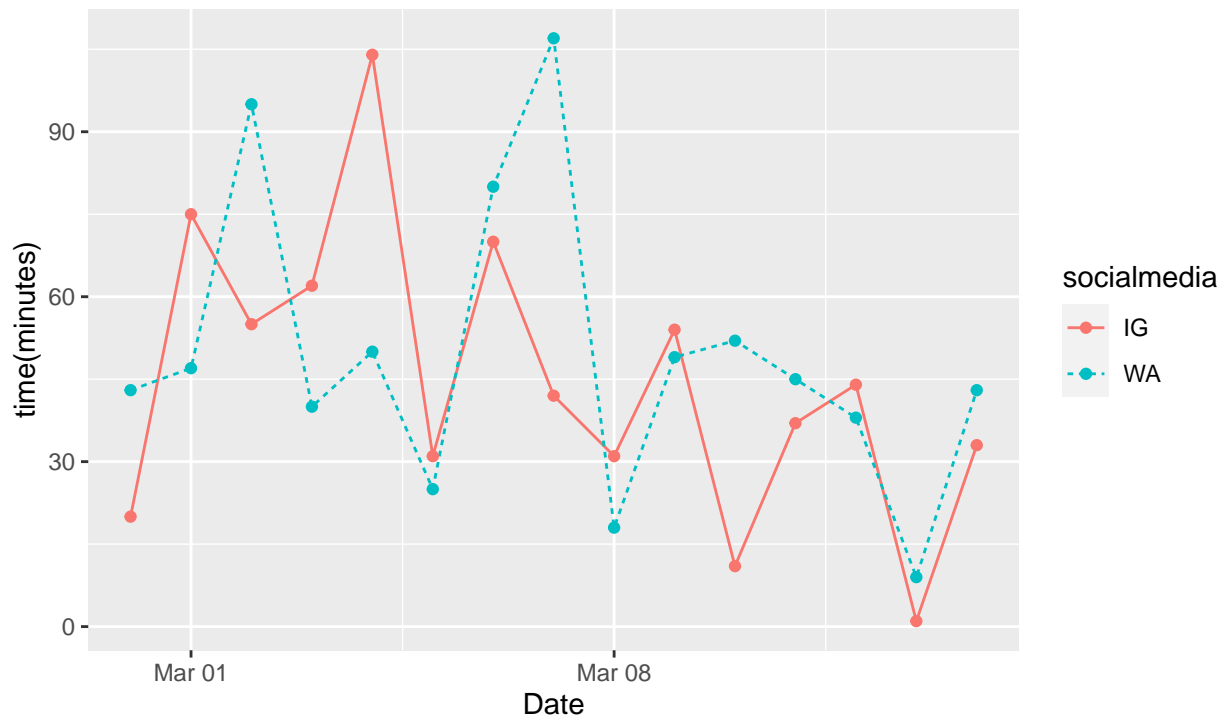
Data visualization 2

The line graph below shows the trend of my usage of social media between 28th February 2021 and 15th March 2021.

```
social_media2<- social_media1%>% #rename variable summary correct
  rename(socialmedia = summary_correct)

#Line graph
p<-ggplot(social_media2, aes(x=date,
                             y=length_hour,
                             linetype=socialmedia)) +
  labs(title = "How my time is split between Instagram
               \n and WhatsApp per day",
        y="time(minutes)",
        x="Date") +
  geom_line(aes(color=socialmedia))+
  geom_point(aes(color=socialmedia))
p
```

How my time is split between Instagram
and WhatsApp per day



Summary table of time spent on my activities

The table shows the average time in minutes spent in each activity, and groups the activities by weekday and weekend.

```
#working with combined dataset
total1 <- total %>% #rename variable
  rename( activity = summary_correct)
#Table
total2 <- total1 %>%
  group_by(type_day,activity)%>%
  summarise(avg_min = mean(length_hour))%>%
  spread(key=type_day, value=avg_min,fill = 1)
```

'summarise()' has grouped output by 'type_day'. You can override using the '.groups' argument.

```
total2
```

```
## # A tibble: 7 x 3
##   activity Weekday Weekend
##   <chr>      <dbl>   <dbl>
## 1 A-DS      80.5     32.5
## 2 A-EC      88.2      94
## 3 A-MB      37.7       1
## 4 IG       50.4     33.2
## 5 R-DS      49.5    159.
## 6 R-MB      52.3       1
## 7 WA       45.9     56.4
```

Summary of my findings:

1. I spend more time on average studying Data Science than any other subject in the 15 days. I spent a total of 1928 minutes doing Data Science reading and assignments, an average of 2.1 hours per day. I spent a total of 270 minutes, an average of 18 minutes per day in Money and Banking course. For Econometrics, I spent a total of 717 minutes for assignments, an average of 47.8 minutes per day. The large amount of time spent in data science could be explained by the amount of time spent debugging the r-codes.
2. In general, I spent more time on WhatsApp (741 minutes) than on Instagram(670) in the 15 days. That is an average 44.6 minutes per day on Instagram and 49.4 minutes per day on WhatsApp. I initially thought that the time I spend between the two apps would be large enough to distinguish, but the difference wasn't large enough, and would require hypothesis testing to confirm if the difference was indeed statistically significant.
3. I expected the amount of time spent on my courses to be less during the weekends than on weekdays. The results were consistent with this assumption. The unusually high amount of time spent on Data Science during the weekend could be explained by the assignment that is often due on Monday.
4. The amount of time I spent on social media on a given day decreased after my first data entry session around 6th of March, which could be explained by my change in behavior because of being recorded.

Reflection

In my data analysis, figuring out the best graphs to represent my findings was challenging. I ended up having to omit some of the variables I had collected to prevent crowding my graphs and making them ineffectual. The challenges I faced in the data collection process were accurately capturing the time I spent on my courses and the behavioral change in my social media usage. When collecting data for the time I spent in my courses outside class, I often took random small breaks that were hard to account for. This resulted in some activities appearing to take more time than they actually did. For my social media usage, I unconsciously changed my behavior since I knew it was being tracked. I probably would have spent more time on social media than what I recorded because of the tracking.

Being both the subject of the experiment and the one who analyses the data, my bias affected the data collection process. It is better to not be the subject if I am the one analyzing the data. I should also be more aware of other people's changes due to being recorded, and factor that into my analysis.

To answer my question of how much time I spend on social media apps, I should collect data over a long period of time, probably over a year. This will make it easier to notice any change in behavior over a longer period of time and the different events that cause the change. Since data will be collected over a long period of time, behavior change due to recording of my time spent on social media will be randomly distributed, making it less of a concern when analyzing data. The data will be manually entered into a spreadsheet from my phone, and I anticipate that this will be time-consuming and tiring.

To answer how much time I spend in my courses outside of class, collecting data at least worth half of the semester will be able to yield more accurate results and show the trends over the days more clearly. The data collection process will be less time-consuming since it is adding time spent in my google calendar. However, there might be a risk of forgetting to do data entry on the calendar.

As someone who provides data, I expect the analysis to be done without bias and my privacy to not be compromised. The entity collecting my data should be transparent and honest about the usage of my data. Under no circumstance should my data be used to reinforce harmful stereotypes.

As someone who analyzes data, I have an ethical responsibility to clearly inform the participants what their data will be used for so that they can make an informed decision when giving out consent. The data collected should always be treated with privacy, and not shared unethically to other people for a different purpose than the intended one. The identity of the participants should never be shared without their knowledge and consent. During analysis, I should never fabricate available data to satisfy a specific objective. Instead, my conclusions should be derived from the result of the analysis done. I also have a moral obligation to ensure that my analysis will not be misused for harmful purposes.