

Week 14

R Studio API Code

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

Libraries

```
library(tidyverse)
library(RMariaDB)
library(caret)
```

Data import and cleaning

SQL code: sql_tbl

```
con <- dbConnect(MariaDB(),
  user      = "rnlander_8960r",
  password  = "rTCo.4vQ2vc-",
  port      = "3306",
  host      = "tntlab.org",
  dbname    = "rnlander_8960")

# see all table names in the database
tbl <- dbGetQuery(con, "SELECT * FROM INFORMATION_SCHEMA.TABLES WHERE TABLE_TYPE = 'BASE TABLE'")
# three tables: demos, responses, socialmedia

# check socialmedia table: "code",
# check responses table: "smu_code", "ident"
# check demos table: "participant_num"
# ident from responses match with participant_num from demos

# join demos and responses by participant number, and join socialmedia by code
sql_tbl <- dbGetQuery(con,
  "SELECT * FROM responses
  INNER JOIN demos
  ON responses.ident = demos.participant_num
  LEFT JOIN socialmedia
  ON responses.smu_code = socialmedia.code")
```

Tidyverse: tidy_tbl

```
demos_tbl <- dbGetQuery(con, "SELECT * FROM demos")
sm_tbl <- dbGetQuery(con, "SELECT * FROM socialmedia")
res_tbl <- dbGetQuery(con, "SELECT * FROM responses")

tidy_tbl <- res_tbl %>%
  inner_join(demos_tbl, by = c("ident" = "participant_num")) %>%
  left_join(sm_tbl, by = c("smu_code" = "code"))
```

Cleaning

```
# mean privacy score, age, and number of social media platforms.
clean_tbl <- tidy_tbl %>%
  # transform variables to numeric
  mutate(rec_events = case_when(rec_events == "NA" ~ NA_integer_,
                                rec_events == "Refused" ~ NA_integer_,
                                rec_events == "Not acceptable at all" ~ 1L,
                                rec_events == "Not very acceptable" ~ 2L,
                                rec_events == "Somewhat acceptable" ~ 3L,
                                rec_events == "Very acceptable" ~ 4L)) %>%
  mutate(rec_products = case_when(rec_products == "NA" ~ NA_integer_,
                                   rec_products == "Refused" ~ NA_integer_,
                                   rec_products == "Not acceptable at all" ~ 1L,
                                   rec_products == "Not very acceptable" ~ 2L,
                                   rec_products == "Somewhat acceptable" ~ 3L,
                                   rec_products == "Very acceptable" ~ 4L)) %>%
  mutate(rec_friends = case_when(rec_friends == "NA" ~ NA_integer_,
                                  rec_friends == "Refused" ~ NA_integer_,
                                  rec_friends == "Not acceptable at all" ~ 1L,
                                  rec_friends == "Not very acceptable" ~ 2L,
                                  rec_friends == "Somewhat acceptable" ~ 3L,
                                  rec_friends == "Very acceptable" ~ 4L)) %>%
  mutate(rec_policial = case_when(rec_policial == "NA" ~ NA_integer_,
                                   rec_policial == "Refused" ~ NA_integer_,
                                   rec_policial == "Not acceptable at all" ~ 1L,
                                   rec_policial == "Not very acceptable" ~ 2L,
                                   rec_policial == "Somewhat acceptable" ~ 3L,
                                   rec_policial == "Very acceptable" ~ 4L)) %>%
  mutate(facebook = case_when(facebook == "Facebook" ~ 1L,
                              facebook != "Facebook" ~ 0L),
         twitter = case_when(twitter == "Twitter" ~ 1L,
                              twitter != "Twitter" ~ 0L),
         instagram = case_when(instagram == "Instagram" ~ 1L,
                                instagram != "Instagram" ~ 0L),
         youtube = case_when(youtube == "Youtube" ~ 1L,
                              youtube != "Youtube" ~ 0L),
         snapchat = case_when(snapchat == "Snapchat" ~ 1L,
                              snapchat != "Snapchat" ~ 0L),
         other = case_when(other == "Other" ~ 1L,
                           other != "Other" ~ 0L)) %>%
  # transform age to factor variable
```

```
mutate(age = factor(age)) %>%
rowwise() %>%
# mean privacy score (mean_score) and number of social media platforms (smp_num)
mutate(mean_score = mean(c(rec_events,rec_products, rec_friends,rec_policial), na.rm=T),
      smp_num = sum(c(facebook, twitter, instagram, other))) %>%
# select necessary variables for analysis
select(mean_score, age, smp_num)
```

Analysis

```
lm_mod <- lm(mean_score ~ age*smp_num,
             data = clean_tbl)
summary(lm_mod)
```

```
##
## Call:
## lm(formula = mean_score ~ age * smp_num, data = clean_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.83044 -0.42876  0.07124  0.47883  1.82124
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.45247    0.07255  33.805 < 2e-16 ***
## age30-49         -0.03439    0.08336  -0.413  0.679973
## age50-64         -0.22767    0.08102  -2.810  0.004973 **
## age65+          -0.41932    0.08265  -5.074  4.07e-07 ***
## ageNA           -1.13997    0.69277  -1.646  0.099933 .
## smp_num           0.11860    0.03511   3.378  0.000735 ***
## age30-49:smp_num -0.01551    0.04099  -0.379  0.705068
## age50-64:smp_num -0.01893    0.04158  -0.455  0.648916
## age65+:smp_num   0.02701    0.04529   0.596  0.550950
## ageNA:smp_num    0.25640    0.29586   0.867  0.386207
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7196 on 4295 degrees of freedom
## (287 observations deleted due to missingness)
## Multiple R-squared:  0.07115,    Adjusted R-squared:  0.0692
## F-statistic: 36.55 on 9 and 4295 DF,  p-value: < 2.2e-16
```

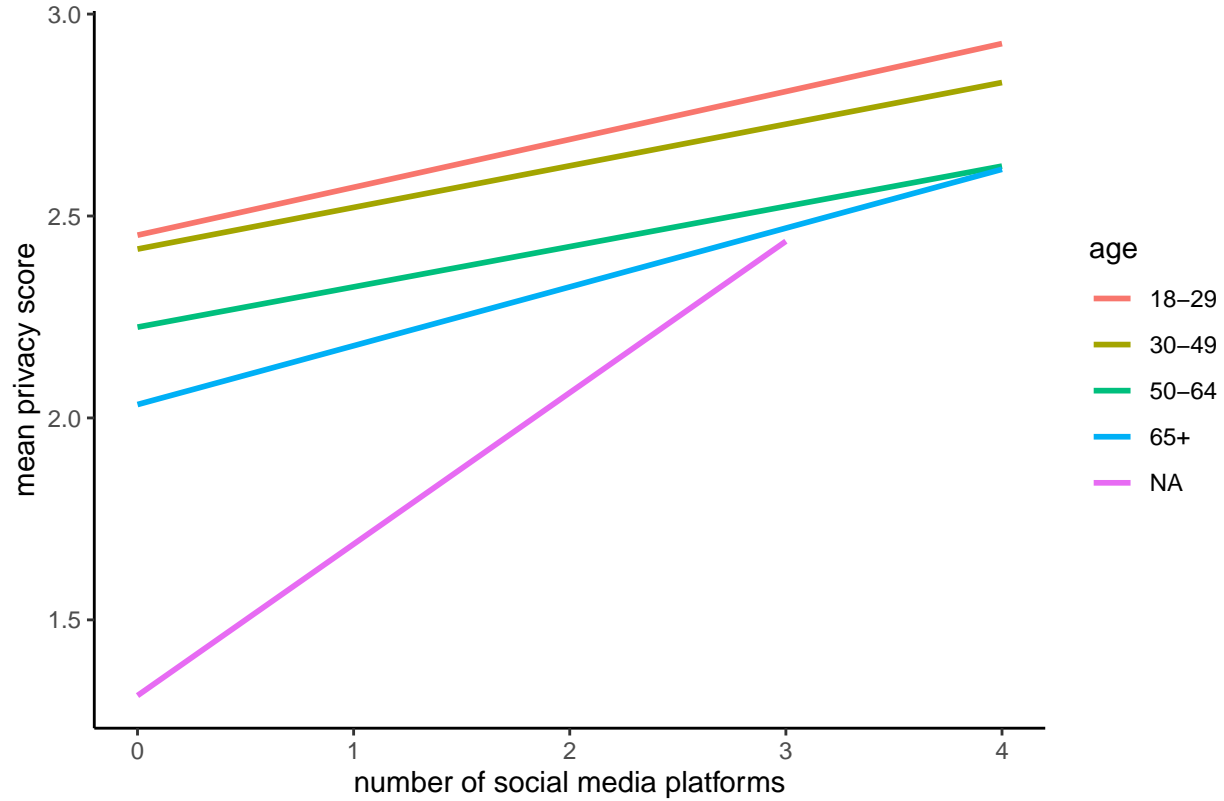
Visualization

```
clean_tbl %>%
ggplot(aes(x = smp_num, y = mean_score, color = age)) +
geom_smooth(method = lm, se = FALSE) +
theme_classic() +
labs(title = "Mean privacy score and Number of social media platforms used by age group",
```

```
x = "number of social media platforms",  
y = "mean privacy score")
```

```
## Warning: Removed 287 rows containing non-finite values (stat_smooth).
```

Mean privacy score and Number of social media platforms used by age gro



Intepretation

People who used more social medial platforms on average had greater acceptance of privacy intrusions. Older people, compared to younger people, had lower acceptance of privacy intrusions. There is no moderation effect of age on the relationship between social media usage and acceptance of privacy intrusions.