R Notebook

R Studio API Code

Libraries & Access

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

```
## -- Attaching packages -----------------------------------------------------------------

## v ggplot2 3.2.1      v purrr   0.3.3
## v tibble  2.1.3      v dplyr   0.8.4
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

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

Data Import and Cleaning

```
con <- dbConnect(MariaDB(),
                 user="rnlander_8960r",
                 password="rTCo.4vQ2vc-",
                 host="tntlab.org")
# dbGetQuery(con,"SHOW DATABASES") # discovered there are two databases
# dbExecute(con,"USE information_schema") # take a look at the first database
# dbGetQuery(con,"SHOW TABLES") # discovered a bunch of irrelevant tables--probably not the correct dat
dbExecute(con,"USE rnlander_8960") # take a look at the second database
```

```
## [1] 0
```

```
dbGetQuery(con,"SHOW TABLES") # yup looks like it
```

```
##   Tables_in_rnlander_8960
## 1                   demos
## 2               responses
## 3              socialmedia
```

```r
# Import Data Separately and Combine using Tidyverse
demos <- dbGetQuery(con,"SELECT * FROM demos")
responses <- dbGetQuery(con,"SELECT * FROM responses")
socialmedia <- dbGetQuery(con,"SELECT * FROM socialmedia")
tidy_tbl <- demos %>%
  left_join(responses,by=c("participant_num"="ident")) %>% # only include if exists in the demos table
  left_join(socialmedia,by=c("smu_code"="code"))

# Import and Combine Data using SQL
sql_tbl <- dbGetQuery(con,"SELECT participant_num,age,sex,device,smu_code,rec_events,rec_products,rec_f
                          FROM demos AS d
                            LEFT JOIN responses AS r
                              ON d.participant_num = r.ident
                            LEFT JOIN socialmedia AS s
                              ON r.smu_code=s.code;") # selected all columns except the two duplicating

# Clean tidy_tbl
tidy_tbl <- tidy_tbl %>%
  mutate_at(vars(matches("rec_")),function(x) case_when(x=="Not acceptable at all"~1,
                                                        x=="Not very acceptable"~2,
                                                        x=="Somewhat acceptable"~3,
                                                        x=="Very acceptable"~4)) %>% # recode responses
  mutate_at(vars(facebook:other),function(x) ifelse(x=="Not selected",0,1)) %>% # recode social media va
  mutate(age=na_if(age,"NA")) %>% # recode odd cases of "NA" in age variable into actual missing value
  mutate(privacy=rowMeans(select(.,rec_events:rec_policial),na.rm=T), # compute variables needed for an
         sm_num=rowSums(select(.,facebook:other)))
```

Analysis

```r
mod1 <- lm(privacy~sm_num,tidy_tbl)
summary(mod1)
```

```
##
## Call:
## lm(formula = privacy ~ sm_num, data = tidy_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.87570 -0.50293 -0.00144  0.49707  1.74559
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.130156   0.023683   89.95   <2e-16 ***
## sm_num      0.124257   0.008887   13.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7296 on 4303 degrees of freedom
##   (288 observations deleted due to missingness)
## Multiple R-squared:  0.04346,    Adjusted R-squared:  0.04323
## F-statistic: 195.5 on 1 and 4303 DF,  p-value: < 2.2e-16
```

```r
mod2 <- lm(privacy~sm_num+age+sm_num*age,tidy_tbl) # age group 18-29 is reference
summary(mod2)
```

```
##
## Call:
## lm(formula = privacy ~ sm_num + age + sm_num * age, data = tidy_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85351 -0.45427  0.04573  0.46927  1.89964
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.410099   0.085863  28.069  < 2e-16 ***
## sm_num           0.080560   0.024518   3.286  0.00103 **
## age30-49        -0.035674   0.097775  -0.365  0.71524
## age50-64        -0.198935   0.095451  -2.084  0.03721 *
## age65+          -0.438068   0.097488  -4.494 7.19e-06 ***
## sm_num:age30-49 -0.000712   0.028985  -0.025  0.98040
## sm_num:age50-64 -0.010992   0.029928  -0.367  0.71342
## sm_num:age65+    0.047767   0.033453   1.428  0.15340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7193 on 4293 degrees of freedom
##   (292 observations deleted due to missingness)
## Multiple R-squared:  0.07163,    Adjusted R-squared:  0.07011
## F-statistic: 47.32 on 7 and 4293 DF,  p-value: < 2.2e-16
```

Model 1 regresses acceptance of privacy intrusions on number of social media platforms used. There is a significantly positive relationship between the two. Model 2 examines the effect of age and finds that compared with the reference group of 18-29, none of the older age groups are significantly different in the relationship between their number of social media platform used and their acceptance of privacy intrusions.

Visualization

```r
ggplot(data=tidy_tbl[!is.na(tidy_tbl$age),],aes(x=sm_num,y=privacy,colour=age))+
  geom_smooth(method="lm")+
  labs(title="Relationship between Number of Social Media Platforms Used and Acceptance of Privacy Intru
       x="Number of Social Media Platforms Used",
       y="Acceptance of Privacy Intrusion")
```

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

Relationship between Number of Social Media Platforms Used and Accept