Ready-To-Analyze Dataset

Ivan, Alex, Hadong

Where our data came from

Our data set was particularly tricky as we had to manually enter data from a survey format from nursing students given to by Professor Kunnen, into a google sheet, in which we converted to a csv and read it in below.

Getting my data and reading it in

Lists of measures in our dataset

1. Participant ID: Unique identifier for each participant
2. Test entry point: When the students took the test, pre, post, and 2 months after using VR goggles (Categorical)
3. Questionnaire Responses: Responses to a series of questions, relating to experience of gaming, and using VR goggles, where choices were A, U, SA, and D, showing the students confidence levels.
4. Age: Age of participant (Probably Categorical?)
5. Gender: Male or female (Binary data one trial)
6. Ethnicity/Race: Ethnicity of participant (Binary data multiple trials)
7. Year of college: What the participants level of education was (Binary data multiple trials)

```
prepost_num <- prepost |>
  # here, write the first and last of the variable names with the Likert ("strongly agree"
  # like first_variable_name : last_variable_name
  # alternately, you can replace "health_assessment:inform_doctor with a list of variable
  # like: c(first_var, second_var, third_var, ..., last_var)
  mutate(across(Health_Assessment:Assesment_Data_Knowledge_Confidence,
                ~case_when(.x == 'SA' ~ 4,
                           .x == 'A' ~ 3,
                           .x == 'U' ~ 2,
```

```
                              .x == 'D' ~ 1,
                              .x == 'SD' ~ 0),
                    # new numeric columns will have names like the original columns,
                    # but with "_num" appended
                    .names = "{.col}_num"),
          knowledge_num = (Health_Assessment_num + Nursing_Process_num + Nursing_Interventi
          confidence_num = (Confidence_of_Health_Assesment_num + Acute_Care_Nursing_Interve
          )

#prepost_num <- prepost_num |>
#  select(Confidence_of_Health_Assesment_num, Actual_Simulation_Expiriences,
#  Year_Of_College, Nursing_Interventions_num, Nursing_Process_num, Gender) |>
#  drop_na()

prepost_num <- prepost_num |>
  mutate(#Actual_Simulation_Expiriences = factor(Actual_Simulation_Expiriences),
         Year_Of_College = factor(Year_Of_College),
         Gender = factor(Gender))

glimpse(prepost_num)
```

```
Rows: 151
Columns: 26
$ Student_ID                                      <dbl> 1, 1, 2, 2, 3, 3, 4, 4~
$ Time_Point                                      <chr> "Pre-test", "Post-test~
$ Health_Assessment                               <chr> "SA", "SA", "U", "U", ~
$ Nursing_Process                                 <chr> "SA", "SA", "A", "U", ~
$ Nursing_Interventions                           <chr> "A", "A", "A", "A", "A~
$ Assesment_Data_Knowledge                        <chr> "U", "A", "U", "A", "A~
$ Confidence_of_Health_Assesment                  <chr> "U", "A", "U", "U", "A~
$ Acute_Care_Nursing_Interventions_Confidence     <chr> "A", "A", "A", "U", "U~
$ Severe_Care_Nursing_Intervention_Confidence     <chr> "U", "A", "A", "U", "U~
$ Assesment_Data_Knowledge_Confidence             <chr> "A", "A", "A", "U", "U~
$ Year_Of_College                                 <fct> J3, NA, J3, J3, J3, NA~
$ Previous_Simulation_Expiriences                 <chr> "Not_Similiar", NA, "N~
$ Age                                             <dbl> 20, NA, 20, 20, 20, NA~
$ Gender                                          <fct> Male, NA, Female, Fema~
$ Race_Ethnicity                                  <chr> "Asian_American", NA, ~
$ Actual_Simulation_Expiriences                   <chr> "Reality_Gaming_Exp", ~
$ Health_Assessment_num                           <dbl> 4, 4, 2, 2, 2, 3, 2, 3~
$ Nursing_Process_num                             <dbl> 4, 4, 3, 2, 2, 2, 3, 3~
```

```
$ Nursing_Interventions_num                          <dbl> 3, 3, 3, 3, 3, 3, 3, 3~
$ Assesment_Data_Knowledge_num                       <dbl> 2, 3, 2, 3, 3, 3, 3, 3~
$ Confidence_of_Health_Assesment_num                 <dbl> 2, 3, 2, 2, 3, 3, 2, 2~
$ Acute_Care_Nursing_Interventions_Confidence_num    <dbl> 3, 3, 3, 2, 2, 2, 3, 3~
$ Severe_Care_Nursing_Intervention_Confidence_num    <dbl> 2, 3, 3, 2, 2, 2, 3, 3~
$ Assesment_Data_Knowledge_Confidence_num            <dbl> 3, 3, 3, 2, 2, 2, 3, 3~
$ knowledge_num                                      <dbl> 3.25, 3.50, 2.50, 2.50~
$ confidence_num                                     <dbl> 2.50, 3.00, 2.75, 2.00~
```

```
mosaic::tally(~knowledge_num, data = prepost_num)
```

```
Registered S3 method overwritten by 'mosaic':
  method                           from
  fortify.SpatialPolygonsDataFrame ggplot2
```

```
knowledge_num
 0.5  1.5 1.75    2 2.25  2.5 2.75    3 3.25  3.5 3.75    4 <NA>
   1    1    2    5   13   20   26   36   21   14    3    4    5
```

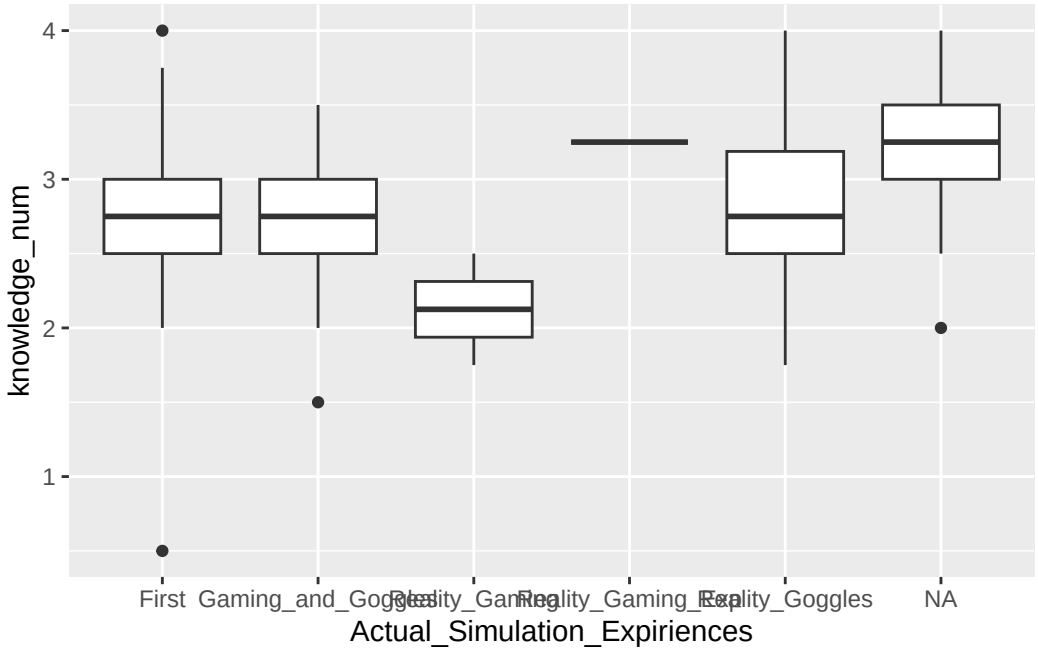
```
mosaic::tally(~confidence_num, data = prepost_num)
```

```
confidence_num
   1 1.25  1.5 1.75    2 2.25  2.5 2.75    3 3.25  3.5 3.75    4
   3    2    4    2    8   18   23   29   41    7    7    2    5
```

```
#str(prepost_num)
```

```
gf_boxplot(knowledge_num ~ Actual_Simulation_Expiriences, data = prepost_num)
```

```
Warning: Removed 5 rows containing non-finite values (`stat_boxplot()`).
```



We chose to use a boxplot plot as this does a good job for the representation of quantitative data points and qualitative distinctions, making it good for distributions of various categories.

The boxplot is showing the different distributions of knowledge of using VR technology compared to different simulation experiences, in which having reality goggles experiences increased the median of knowledge.

```
knowledge_model <- lm( #glmmTMB(
knowledge_num ~ Actual_Simulation_Expiriences +
  Year_Of_College + Gender,
  data = prepost_num,
  #family = poisson(link = 'log')
)
  options(scipen = 999) #Makes the results in non scientific notation

summary(knowledge_model)
```

```
Call:
lm(formula = knowledge_num ~ Actual_Simulation_Expiriences +
    Year_Of_College + Gender, data = prepost_num)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.21790 -0.28155  0.00683  0.28210  1.25683

Coefficients:
                                                Estimate Std. Error t value
(Intercept)                                      2.74317    0.07631  35.948
Actual_Simulation_ExpiriencesGaming_and_Goggles -0.02527    0.11361  -0.222
Actual_Simulation_ExpiriencesReality_Gaming     -0.61817    0.33902  -1.823
Actual_Simulation_ExpiriencesReality_Gaming_Exp  0.65033    0.49676   1.309
Actual_Simulation_ExpiriencesReality_Goggles     0.04333    0.13234   0.327
Year_Of_CollegeJ4                                0.27352    0.20204   1.354
GenderMale                                      -0.14350    0.16548  -0.867
                                                          Pr(>|t|)
(Intercept)                                     <0.0000000000000002 ***
Actual_Simulation_ExpiriencesGaming_and_Goggles              0.8245
Actual_Simulation_ExpiriencesReality_Gaming                  0.0717 .
Actual_Simulation_ExpiriencesReality_Gaming_Exp              0.1939
Actual_Simulation_ExpiriencesReality_Goggles                 0.7442
Year_Of_CollegeJ4                                            0.1793
GenderMale                                                   0.3882
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4671 on 87 degrees of freedom
  (57 observations deleted due to missingness)
Multiple R-squared:  0.07904,   Adjusted R-squared:  0.01553
F-statistic: 1.244 on 6 and 87 DF,  p-value: 0.2918
```

```r
    # Make predictions using the fitted model
   # preds <- predict(knowledge_model, newdata = model)

    # Calculate residuals
    resids <- resid(knowledge_model)

    # Residuals vs. fitted values plot
 #   gf_point(resids ~ preds, data = model)

    # Histogram of residuals
#    gf_histogram(~ resids, data = model, bins = 10)
```

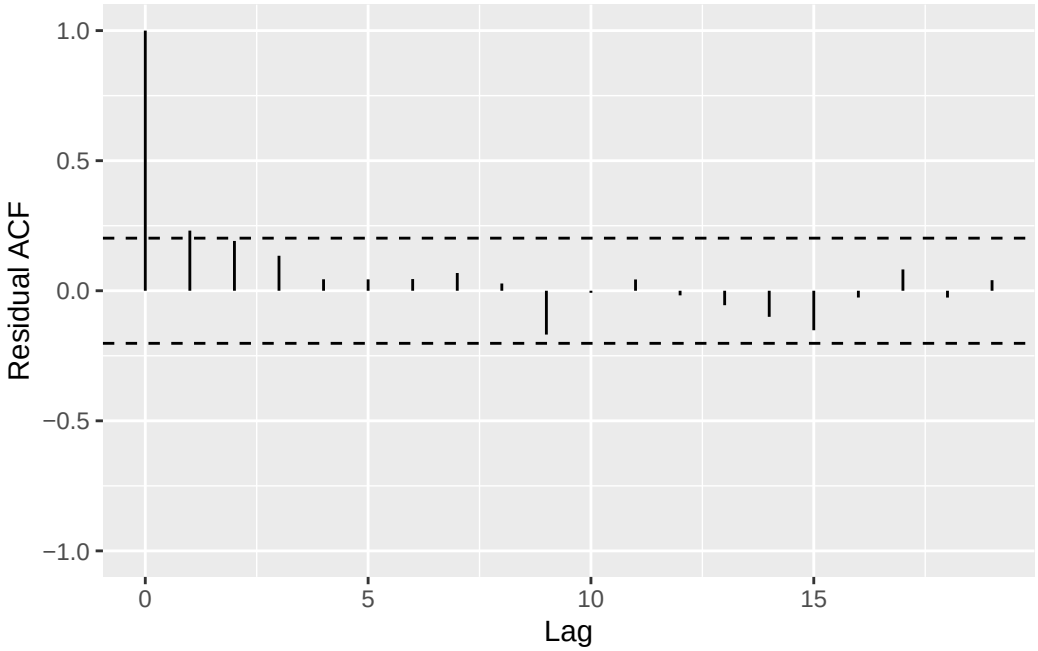
```
fitted(knowledge_model)
```

```
       1        3        4        5        7        9       11       13
3.250000 2.717900 2.743172 2.743172 2.743172 2.717900 2.743172 2.642998
      15       18       21       23       25       28       31       33
2.125000 2.717900 2.599671 2.642998 2.717900 2.743172 2.743172 2.717900
      35       37       39       42       44       46       48       50
2.786499 2.743172 2.743172 2.786499 2.599671 2.743172 2.125000 2.743172
      53       55       56       57       58       59       60       61
2.743172 2.786499 2.786499 2.717900 2.717900 2.717900 2.717900 2.743172
      62       63       64       65       66       67       68       69
2.743172 3.016696 3.016696 2.717900 2.717900 2.717900 2.743172 2.743172
      70       71       72       73       74       75       76       77
2.743172 2.743172 2.717900 2.717900 2.717900 2.717900 2.717900 2.743172
      78       79       80       81       82       83       84       85
2.743172 2.743172 3.016696 3.016696 2.786499 2.786499 2.786499 2.786499
      86       87       88       89       90       91       92       93
2.786499 2.786499 2.786499 2.786499 2.743172 2.743172 2.743172 2.743172
      94       95       96       97       98       99      100      101
2.743172 2.743172 2.717900 2.717900 2.717900 2.717900 2.717900 2.717900
     102      104      106      108      110      116      118      120
2.786499 2.599671 2.743172 2.599671 2.717900 2.916522 2.717900 2.786499
     123      125      127      129      131      134      136      138
2.743172 2.717900 2.743172 2.574399 2.717900 2.717900 2.743172 2.743172
     140      142      144      146      148      150
2.743172 2.743172 2.743172 2.574399 3.016696 2.786499
```

Assessment

Independence Test

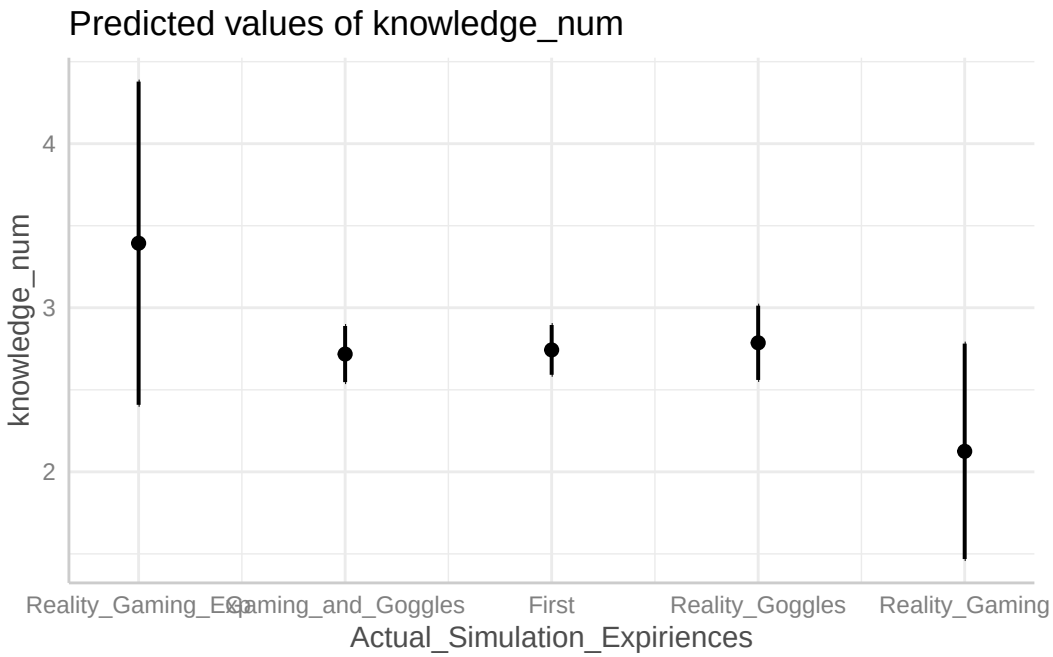
```
s245::gf_acf(~resid(knowledge_model)) |> gf_lims(y = c(-1, 1))
```



Linearity Test

```
#gf_point(log(Confidence_of_Health_Assesment_num) ~ Actual_Simulation_Expiriences, data =
```

```
ggpredict(knowledge_model, terms = "Actual_Simulation_Expiriences") |>
  plot()
```



Explanation

For this week's submission, we were not able to complete all the graphs that we need for our model, however, we implemented your strategy of taking the average of certain variables we are using for a particular question. I also implemented the change to linear regression instead of count regression. There are definitely still some things we need to figure out, especially with Professor Kunnen, to ask her if we are able to merge questions together. From there we are able to make further progress in our graphs. However, we have an idea of where we need to go from here and we included our exploring graph (boxplot), one of our assessment models (ACF) and also our prediction plot. The rest of the graphs we were not able to complete yet, however, once we are able to confirm our next steps with Professor Kunnen, we will be able to make more progress.