

Effect of Video Game on Aggression

Huiting Wu

Sharmeen Kapoorwala

May 6, 2025

Introduction

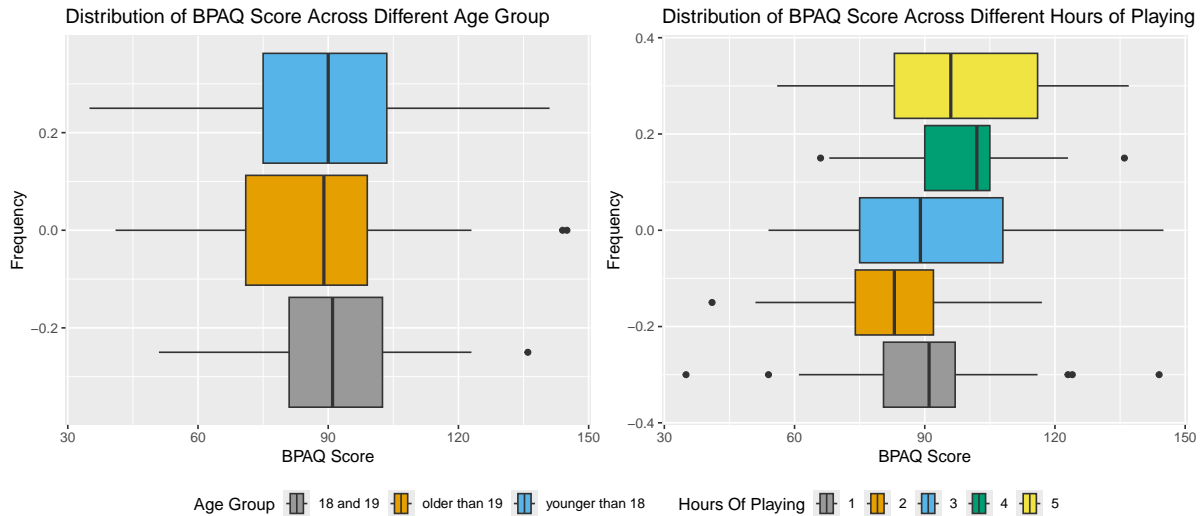
Numerous studies have examined the connection between playing violent video games and acting aggressively, particularly in regard to young people's increased video game consumption. While some studies believe that playing violent video games might be a harmless way to relieve stress, others claim that it can actually increase violence. This dataset was collected in Pakistan in order to study the relationship between gamers' interactions with video games and how such interactions affect their aggressive behavior. The purpose of this analysis is to find trends or correlations that can provide insight into whether playing violent video games is a good indicator of violence. Our analysis's research question is: Do age groups and video game playing time have an impact on a player's level of aggression? The study we conducted is based on the hypothesis $H_0 : \text{Age groups and the amount of time spent playing video games do not significantly affect BPAQ scores. } H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \dots = \alpha\beta_{35} = 0$. $H_1 : \text{Age groups and the amount of time spent playing video games have a substantial impact on BPAQ scores. } H_1 : \alpha\beta_{ij} \neq 0$ where i (age_group) = 1,2,3 and j (hours_of_playing) = 1,2,3,4,5

Methods

This dataset includes a total of 147 observations and 43 variables out of which 1 variable is for timestamp, 5 questions are the background questions including name, age, gender, city, type of family (nuclear,joint), favorite type of game, favorite game, Hours spent on playing and 33 other behavioral questions. The questions were answered in the format of likert scale ("Strongly disagree", "Disagree", "Neither agree nor disagree", "Strongly agree", "Agree") so they were converted into numeric value of ("1", "2", "3", "4", "5") and hours of playing were initially answered as ("less than 1 hour", "more than 1 hour", "more than 2 hour", "more than 3 hour", "more than 5 hour") and they were converted into numeric value of ("1", "2", "3", "4", "5"). A new variable was added in dataset names as BPAQ Score and it stores the total count of questions that were answered. The higher number shows the higher effect on behavior (aggression) after playing video games. The dataset analyzes the correlation between video game play and aggression in people between the ages of 15 and 32. The main goal is to determine whether aggression scores are affected by age group and video game playing hours. Once preprocessing has been completed (data cleaning, encoding, BPAQ score), the analysis was further proceeded by data visualizations (box plots and histograms for hours of playing and age groups, interaction plot) , two-way ANOVA for fixed effects (for interaction between age groups and hours of playing, assumptions checked (plots, leveneTest and Shapiro-Wilk Test) and then Post-Hoc Analysis (Tukey pairwise comparisons).

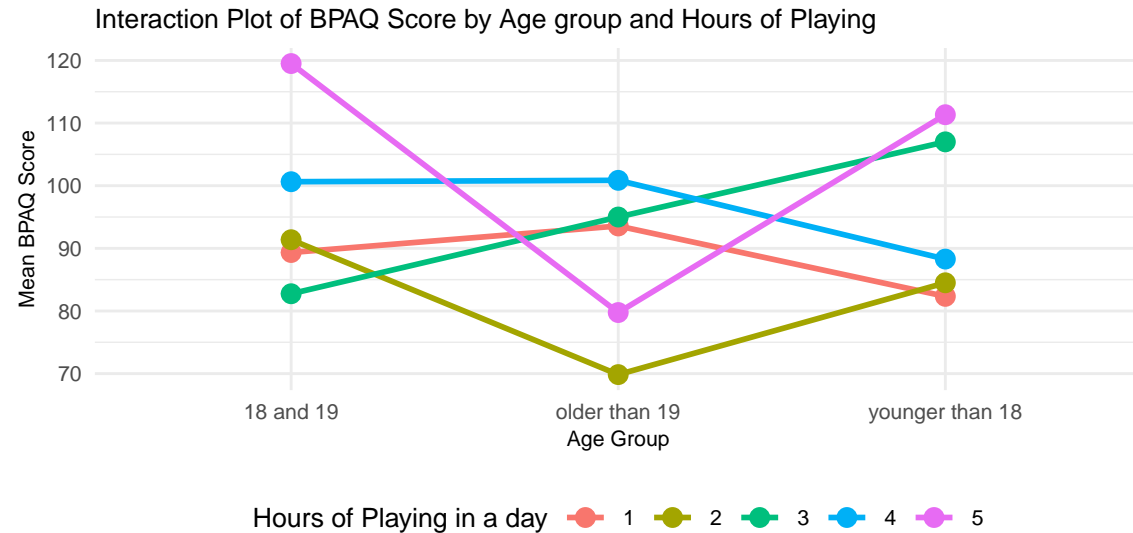
Results

Data Visualization



The boxplot distribution of the BPAQ scores among hours of playing show that the median of the scores is different. There may be a mean score difference among hours of playing but the box plot of the BPAQ scores among different age group shows that the median of the scores is similar so there may not be significant mean score difference among age groups.

Interaction Plot



The interaction plot shows that there is an interaction effect among some combination of age group and hours of playing. There is an interaction effect on the BPAQ score.

Two-way ANOVA for fixed effects

- Check Assumptions

Assumptions Check (Interaction Effects Model)

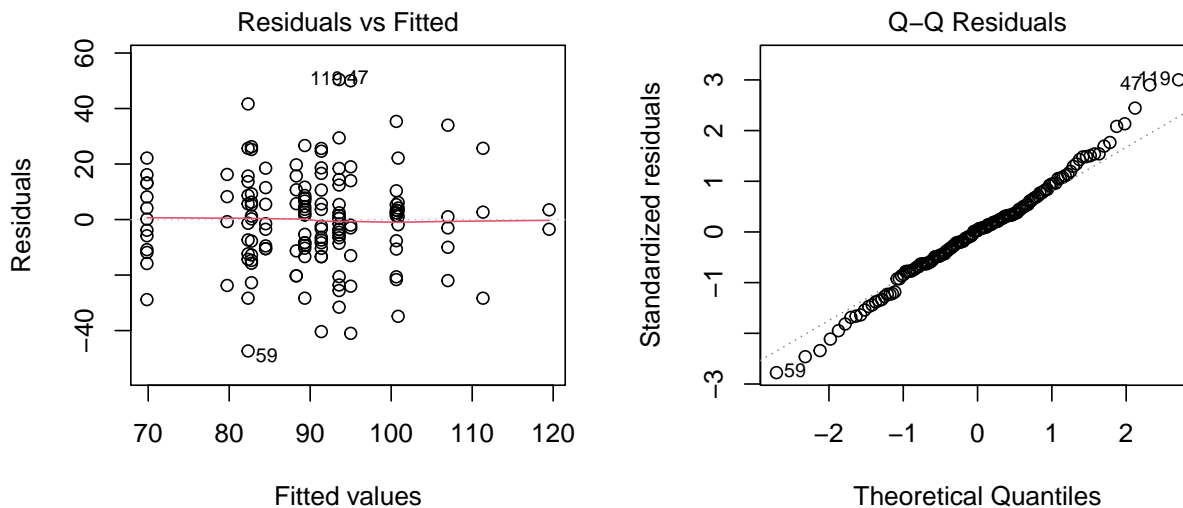


Table 1: Levene's Test

	Df	F value	Pr(>F)
group	14	0.8702815	0.592274
	132	NA	NA

Table 2: Shapiro-Wilk test

	Statistic	p_value
W	0.9897172	0.356806

The variances between the groups are roughly identical, according to the residuals vs. fitted plot. Performed a Levene test, which validates the equal variance assumption with a p-value of 0.5923. The usual QQ line is followed by the majority of the points in the normal QQ plot and the Shapiro-Wilk test result, which showed a p-value of 0.3568. The assumption of normalcy is met.

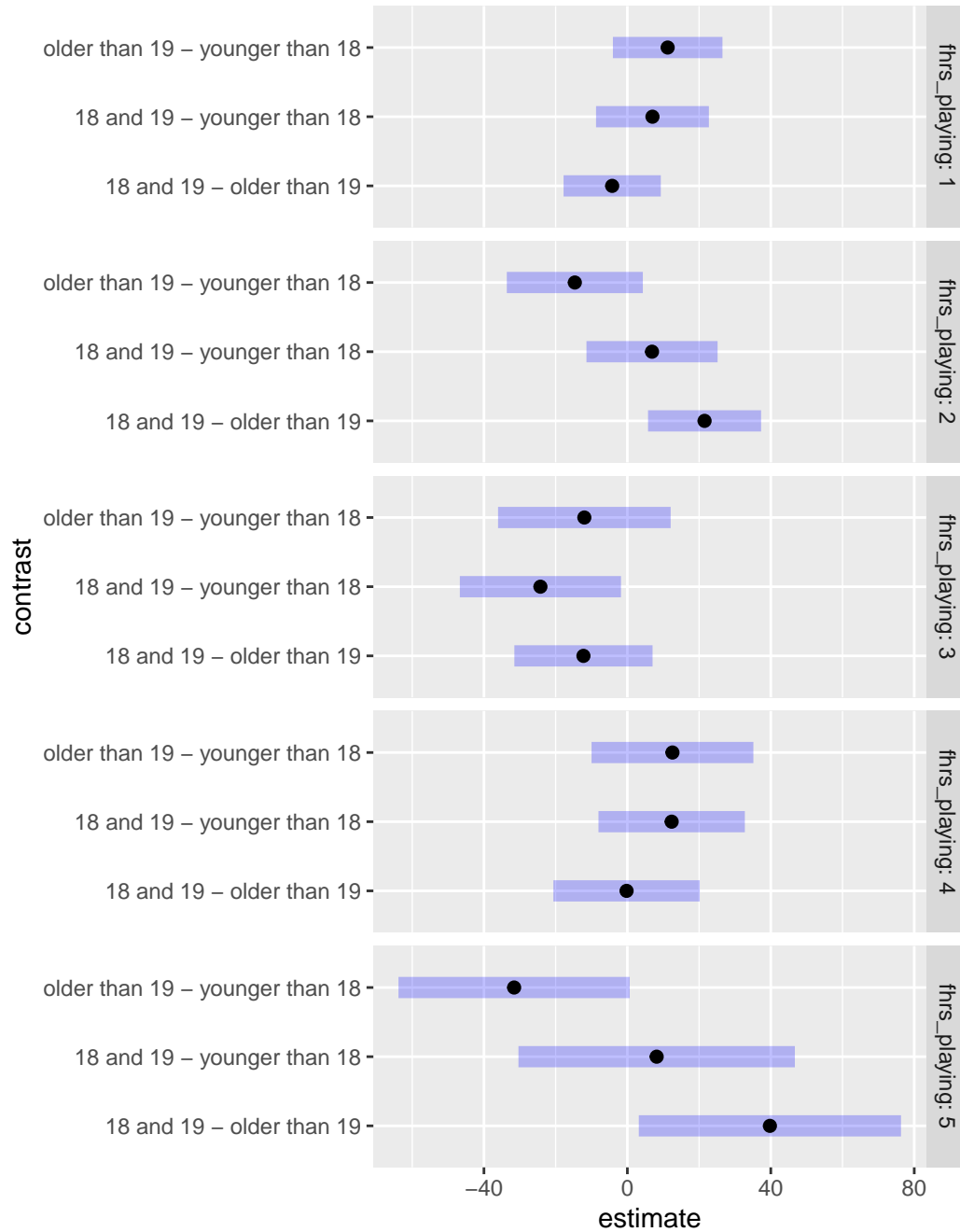
- Anova Table

Table 3: Type II ANOVA Table

	Sum Sq	Df	F value	Pr(>F)
Age Group	430.1952	2	0.67812	0.50933
Hours of Playing	4339.0619	4	3.41983	0.01073
Interaction (Age \times Hours)	9683.8620	8	3.81616	0.00046
Residuals	41870.2497	132	NA	NA

The interaction effect is significant. The effect of age group and hours of playing are affecting the mean BPAQ score interactively.

- Post-Hoc Analysis



There is no significant difference in mean scores for those playing for less than 1 hour and less

than 3 hours among age groups. The 18–19 year-old show the highest aggression overall, and those over the age of 19 are the least affected possibly because they can manage their emotions better or play different types of games.

Discussion

Limitations Since the data is self-reported, there may be inaccuracies due to participants not reporting their behavior honestly. Also, other factors like game type and home life might influence aggression, which could affect the results. Unbalanced dataset might have an impact on the significance.

Implications We should give age-specific advice about gaming time and aggression, as its effects vary by age and hours of playing. Also, the significant interaction raises the possibility that younger people are more vulnerable to increased time spent in playing video games more frequently. Other factors like the kind of game, whether someone plays alone or with others, it might also influence aggression and should be explored further. These results could direct teen-focused digital health awareness, educational policies, and supervision by parents.

References

- Muhammad Shamoeel Ul Naeem, Atiq Ur Rehman, Zuhair Khan Barozai, and Zeshan Nazir. (2024). Effects of Video Games on Aggression [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/10201632>
- Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. <https://doi.org/10.1037/0022-3514.63.3.452>
- Wikipedia contributors. (2024, November 6). Buss–Perry aggression questionnaire. Wikipedia. https://en.wikipedia.org/wiki/Buss%E2%80%93Perry_Aggression_Questionnaire
- Gamma. (n.d.). Gamma – Create beautiful presentations in minutes. Retrieved May 4, 2025, from <https://gamma.app/>
- Pedersen T (2025). patchwork: The Composer of Plots. R package version 1.3.0.9000, <https://github.com/thomasp85/patchwork>, <https://patchwork.data-imagist.com>.
- Allaire, JJ, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2024. Rmarkdown: Dynamic Documents for r. <https://github.com/rstudio/rmarkdown>.
- Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. R Markdown: The Definitive Guide. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.

Appendix

Complete R code with comments Additional plots and diagnostic tests Raw data summary statistics

```
library(dplyr)
library(stringr)
library(car)
library(phia)
library(knitr)
```

Import Data

```
game <- read.csv("Effects of Violent Video Games On Aggression CSV MSDOS.csv", stringsAsFactors = FALSE)
```

Data cleaning

- Rename the variable names

```
fgame <- game |>
  mutate(across(where(is.character), as.factor)) |>
  rename(age = What.is.your.age.,
         family_type = Type.of.Family,
         fav_game = Name.the.video.game.you.usually.play,
         violent_game_hr = How.much.time.do.you.play..violent..video.games.specifically.,
         city = City..Residencial.status,
         game_type = What.type.of.video.games.do.you.typically.play.,
         behavioral_chng = What.changes.on.behaviour.have.you.experienced.in.yourself.after.,
         hot_head = Some.of.my.Friends.think.I.am.hothead,
         hours_of_playing = How.many.hours.do.you.play.Video.Games.in..a.day.,
         violence = If.I.have.to.resort.to.violence.to.protect.my.rights..I.will,
         self_doubt = When.people.are.especially.nice.to.me..I.wonder.what.they.want,
         conflict = I.tell.my.friends.openly.when.I.disagree.with.them,
         mad = Once.I.became.so.mad..I.broke.things,
         argue = When.people.disagree.with.me.I.get.into.arguments,
         abusive = When.I.argue.I.use.abusive.language,
         hot_temper = I.am.a.hot.tempered.person,
         suspicious = I.am.suspicious.of.strangers.who.are.too.friendly,
         threaten = I.have.threatened.some.people.whom.I.know,
```

```

anger = I.can.get.angry.suddenly.but.get.over.it.quickly,
hit = If.I.am.provoked.enough..I.will.hit.another.person,
no_fear = When.people.annoy.me..I.openly.tell.them.what.I.think.of.them,
hit_noreason = I.may.hit.a.person.for.no.good.reason,
unfair = Sometimes.I.feel.I.am.being.treated.unfairly,
control = I.have.trouble.controlling.my.temper,
irritate = When.frustrated..I.show.my.irritation,
laugh = Sometimes.I.feel.people.are.laughing.behind.my.back,
disagree= I.often.disagree.with.people,
hit_back = If.somebody.hits.me..I.hit.back,
explode = I.sometimes.feel.like.exploding.for.no.good.reason,
opportunity = I.feel.other.people.always.take.over.the.opportunity.and.I.miss.it,
fight= there.are.people.who.pushed.me.so.far.that.we.begin.fighting,
talk_back = I.know.that.some.so.called..friends..talk.about.me.behind.my.back,
argumentive = My.friends.say.that.I.am.a.bit.argumentative,
lose_temp = Sometimes.I.lose.temper.for.no.good.reason,
fighter = I.get.into.fights.a.little.more.than.a.normal.person,
urge_to_hit = sometimes.I.can.not.control.the.feeling.to.hit.another.person,
jealous = I.sometimes.get.too.much.jealous.from.people,
bitter = I.dont.know.why.sometimes.I.feel.bitter.about.things,
deli_behave = Have.you.ever.been.involved.in.delinquent.behaviour..like.stealing..b
impact_of_play = Do.you.believe.that.playing.violent.video.games.can.lead.to.aggress
)

```

- Convert the answers into likert scores

```

likert_cols <- c("violence","self_doubt","conflict","mad","argue","abusive","hot_temper",
               "suspicious","threaten","anger","hit","no_fear","hit_noreason","unfair","con
               "laugh","disagree","hit_back","explode","opportunity","fight","talk_back",
               "argumentive","lose_temp","fighter","urge_to_hit","jealous","bitter")

likert_scale <- c("Strongly disagree" = 1,
                 "Disagree" = 2,
                 "Neither agree nor disagree" = 3,
                 "Strongly agree" = 4,
                 "Agree" = 5)

fgame$hot_head <- likert_scale[fgame$hot_head]

for (col in likert_cols){
  if (col %in% names(fgame)) {

```

```

    cleaned_values <- trimws(fgame[[col]])
    converted <- likert_scale[cleaned_values]
    if(length(converted) == nrow(fgame)) {
      fgame[[col]] <- converted
    }
    else
      warning(paste("Conversion Failed: ", col))
  }
}

likert_cols_null <- c("urge_to_hit","jealous","bitter")

likert_scale_null <- c("Strongly disagree" = 1,
  "Disagree" = 2,
  "Neither agree nor disagree" = 3,
  "Strongly agree" = 4,
  "Agree" = 5)

fgame$urge_to_hit <- likert_scale_null[fgame$urge_to_hit]
fgame$jealous <- likert_scale_null[fgame$jealous]
fgame$bitter <- likert_scale_null[fgame$bitter]

# Due with the missing values

fgame[likert_cols_null] <- lapply(fgame[likert_cols_null], function(x){
  x[is.na(x)] <- 0
  return(x)
})

likert_cols_hour <- c(fgame$hours_of_playing,fgame$violent_game_hr)
likert_scale_hour <- c("less than 1 hour" = 1,
  "more than 1 hour" = 2,
  "more than 2 hour" = 3,
  "more than 3 hour" = 4,
  "more than 5 hour" = 5)

fgame$hours_of_playing <- likert_scale_hour[fgame$hours_of_playing]
fgame$violent_game_hr <- likert_scale_hour[fgame$violent_game_hr]

fgame$BPAQ_score <- rowSums(fgame[likert_cols], na.rm = T)

# group the game type and age group

```

```
fgame <- fgame %>% mutate(ftype = case_when(
  game_type == "action" ~ "action",
  game_type == "fighting" ~ "fighting",
  TRUE ~ "others"
)) %>% mutate(
  age_chr = as.character(age),
  age_numeric = str_extract_all(age_chr, "\\d+") %>%
    lapply(as.numeric) %>%
    sapply(function(x) max(x, na.rm = TRUE)),
  age_group = case_when(
    age_numeric < 18 ~ "younger than 18",
    age_numeric %in% c(18, 19) ~ "18 and 19",
    age_numeric > 19 ~ "older than 19")) %>%
  select(-age_numeric, -age_chr)

fgame$fhrs_playing <- as.factor(fgame$hours_of_playing)
levels(fgame$fhrs_playing)
```

```
[1] "1" "2" "3" "4" "5"
```

```
levels(fgame$age_group)
```

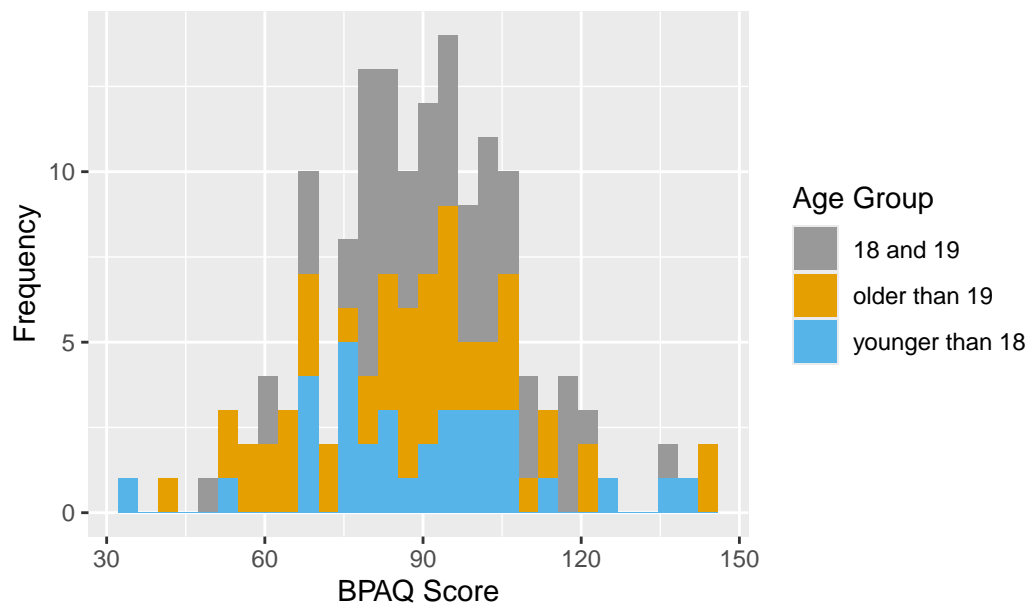
```
NULL
```

Visualization

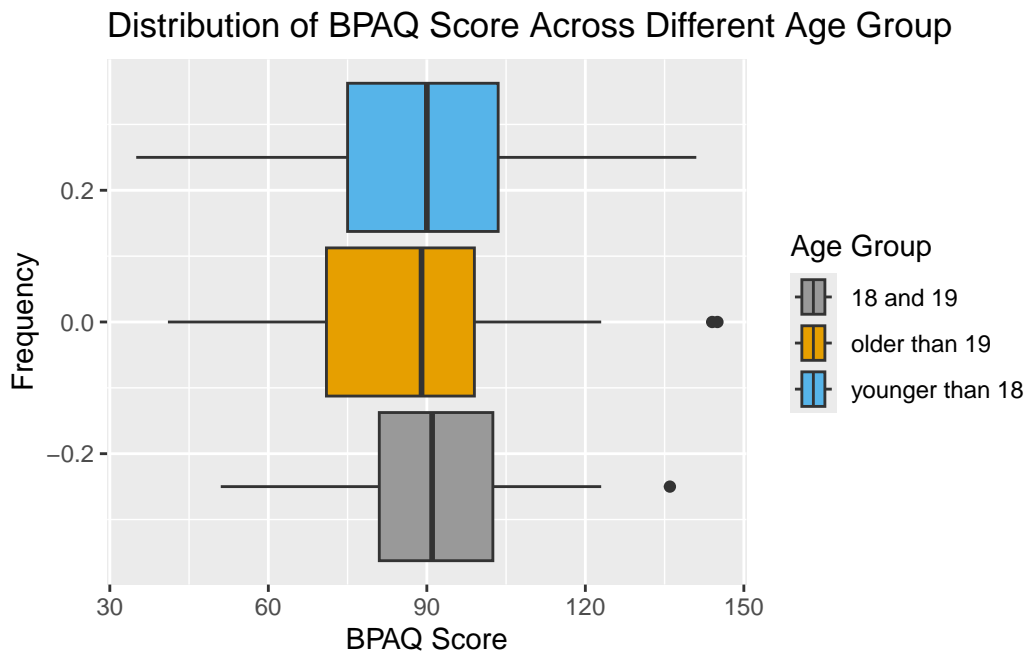
```
fgame %>% ggplot(aes(x = BPAQ_score, fill = age_group)) +
  geom_histogram() +
  scale_fill_manual(values=c("#999999", "#E69F00", "#56B4E9")) +
  labs(x = "BPAQ Score",
       y = "Frequency",
       fill = "Age Group",
       title = "Distribution of BPAQ Score Across Different Age Group")
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Distribution of BPAQ Score Across Different Age Group



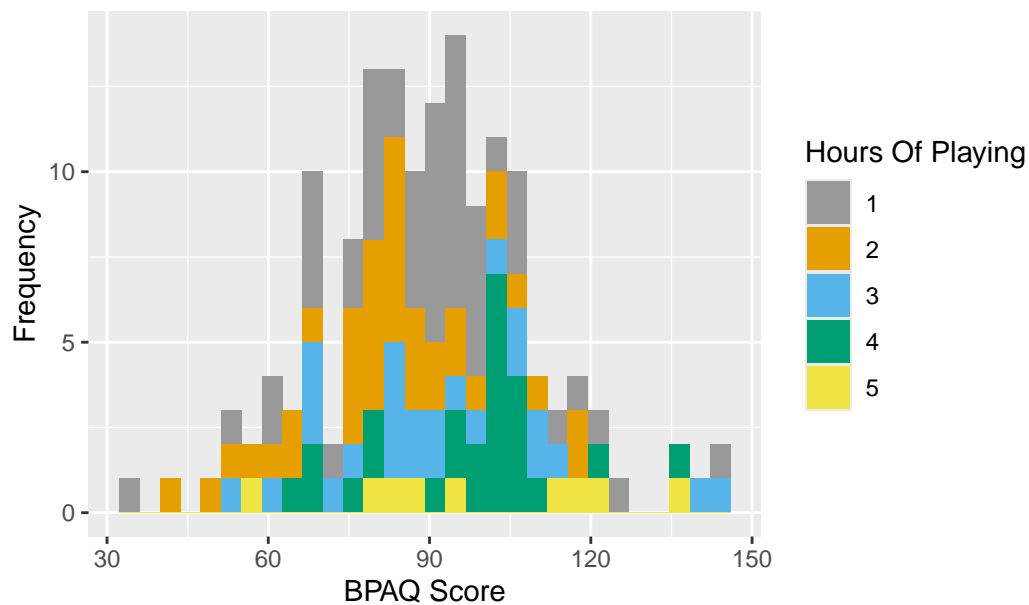
```
fgame %>% ggplot(aes(x = BPAQ_score, fill = age_group)) +
  geom_boxplot() +
  scale_fill_manual(values=c("#999999", "#E69F00", "#56B4E9")) +
  labs(x = "BPAQ Score",
       y = "Frequency",
       fill = "Age Group",
       title = "Distribution of BPAQ Score Across Different Age Group")
```



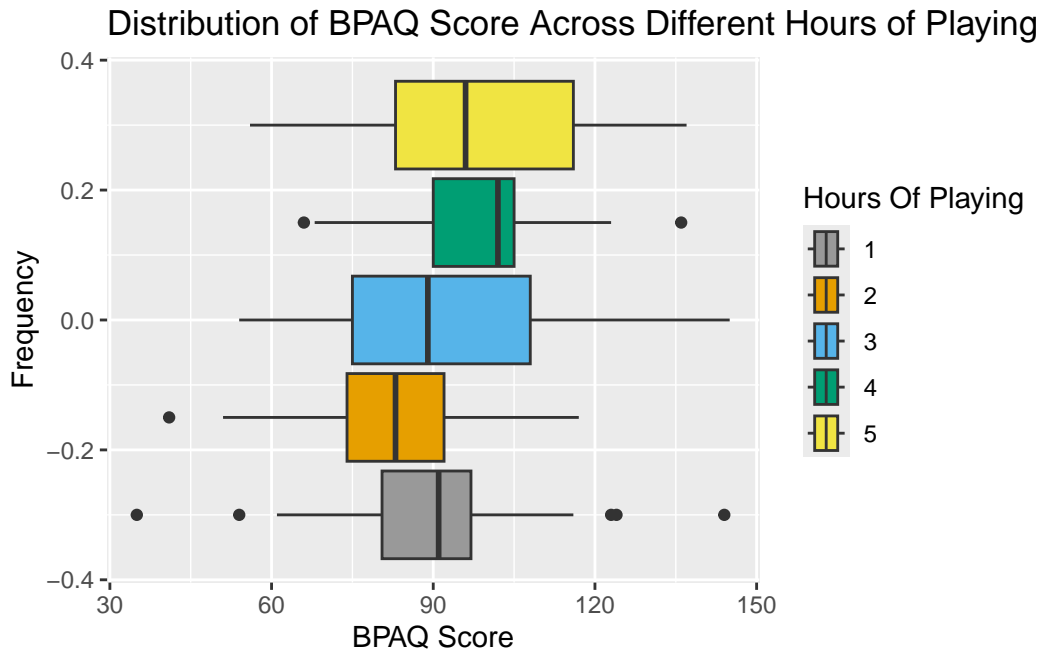
```
fgame %>%
  ggplot(aes(x = BPAQ_score, fill = fhrs_playing)) +
  geom_histogram() +
  scale_fill_manual(values = c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442")) +
  labs(
    x = "BPAQ Score",
    y = "Frequency",
    fill = "Hours Of Playing",
    title = "Distribution of BPAQ Score Across Different Hours of Playing"
  )
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Distribution of BPAQ Score Across Different Hours of Playing



```
fgame %>%
  ggplot(aes(x = BPAQ_score, fill = fhrs_playing)) +
  geom_boxplot() +
  scale_fill_manual(values = c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442")) +
  labs(
    x = "BPAQ Score",
    y = "Frequency",
    fill = "Hours Of Playing",
    title = "Distribution of BPAQ Score Across Different Hours of Playing"
  )
```



Check Balance

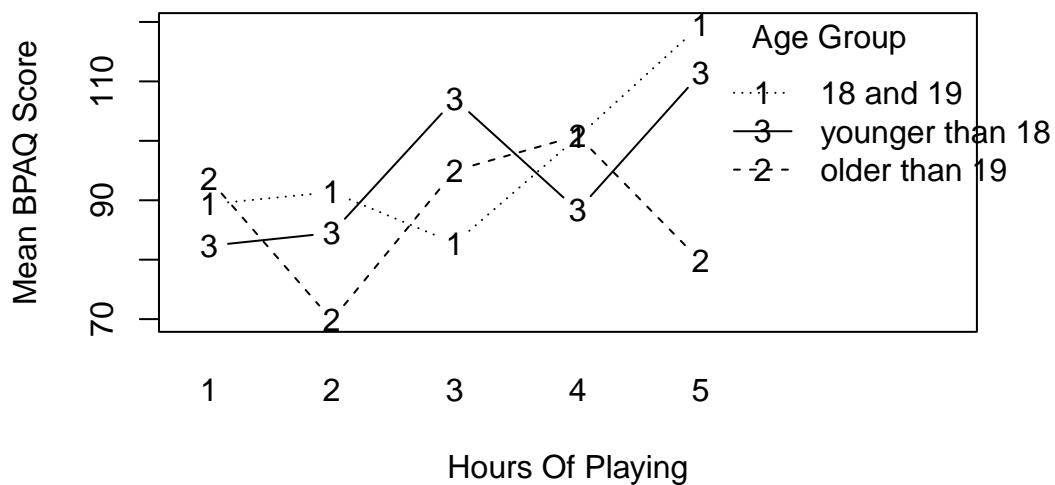
```
fgame %>% group_by(age_group, hours_of_playing) %>% count()
```

```
# A tibble: 15 x 3
# Groups:   age_group, hours_of_playing [15]
  age_group      hours_of_playing      n
  <chr>          <dbl> <int>
1 18 and 19              1      18
2 18 and 19              2      16
3 18 and 19              3      12
4 18 and 19              4      11
5 18 and 19              5       2
6 older than 19          1      21
7 older than 19          2      13
8 older than 19          3       8
9 older than 19          4       7
10 older than 19         5       4
11 younger than 18       1      12
12 younger than 18       2       8
13 younger than 18       3       5
```

14	younger than 18	4	7
15	younger than 18	5	3

Interaction Plot

```
with(fgame, interaction.plot(trace.factor = age_group,
                             x.factor = hours_of_playing,
                             response = BPAQ_score, type = "b",
                             xlab = "Hours Of Playing",
                             ylab = "Mean BPAQ Score",
                             trace.label = "Age Group" ))
```



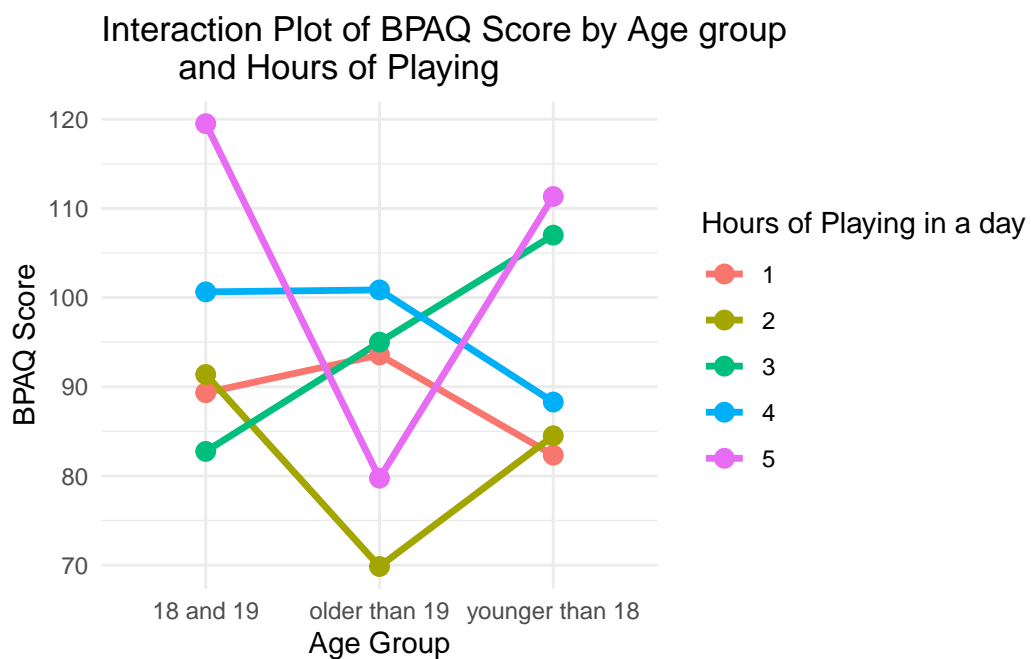
```
fit1 <- aov(BPAQ_score ~ as.factor(age_group)*fhrs_playing, data = fgame)
summary(fit1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(age_group)	2	402	200.9	0.633	0.532373
fhrs_playing	4	4339	1084.8	3.420	0.010735 *
as.factor(age_group):fhrs_playing	8	9684	1210.5	3.816	0.000461 ***
Residuals	132	41870	317.2		

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
plot(interactionMeans(fit1))

ggplot(fgame, aes(x=age_group, y = BPAQ_score, color= fhrs_playing, group = fhrs_playing)) +
  stat_summary(fun = mean, geom = "line" , size = 1.2) +
  stat_summary(fun = mean, geom = "point" , size = 3) +
  labs (title = "Interaction Plot of BPAQ Score by Age group
    and Hours of Playing",
    x = "Age Group",
    y = "BPAQ Score",
    color= "Hours of Playing in a day") +
  theme_minimal()
```



ANOVA Table

```
# test for main effects
main <- lm(BPAQ_score ~ age_group + fhrs_playing, data = fgame)
Anova(main)
```

Anova Table (Type II tests)

Response: BPAQ_score

	Sum Sq	Df	F value	Pr(>F)
age_group	430	2	0.5841	0.55895
fhrs_playing	4339	4	2.9458	0.02248 *
Residuals	51554	140		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
# test for interaction effect
```

```
model <- lm(BPAQ_score ~ age_group*fhrs_playing, data = fgame)
Anova(model)
```

Anova Table (Type II tests)

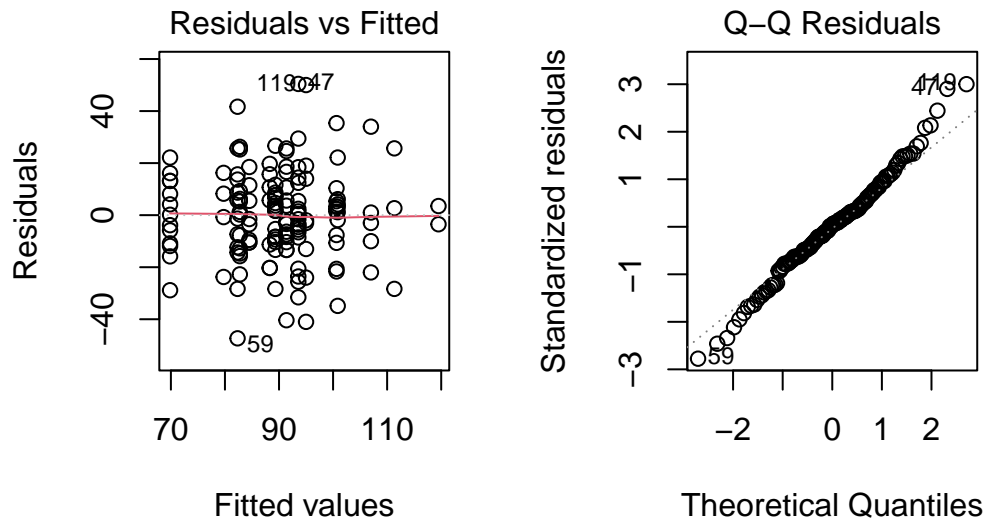
Response: BPAQ_score

	Sum Sq	Df	F value	Pr(>F)
age_group	430	2	0.6781	0.5093317
fhrs_playing	4339	4	3.4198	0.0107348 *
age_group:fhrs_playing	9684	8	3.8162	0.0004606 ***
Residuals	41870	132		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Check Assumptions

```
# interaction effect models
par(mfrow = c(1,2))
plot(model, 1:2)
```



```
leveneTest(BPAQ_score~ age_group* fhrs_playing, data = fgame)
```

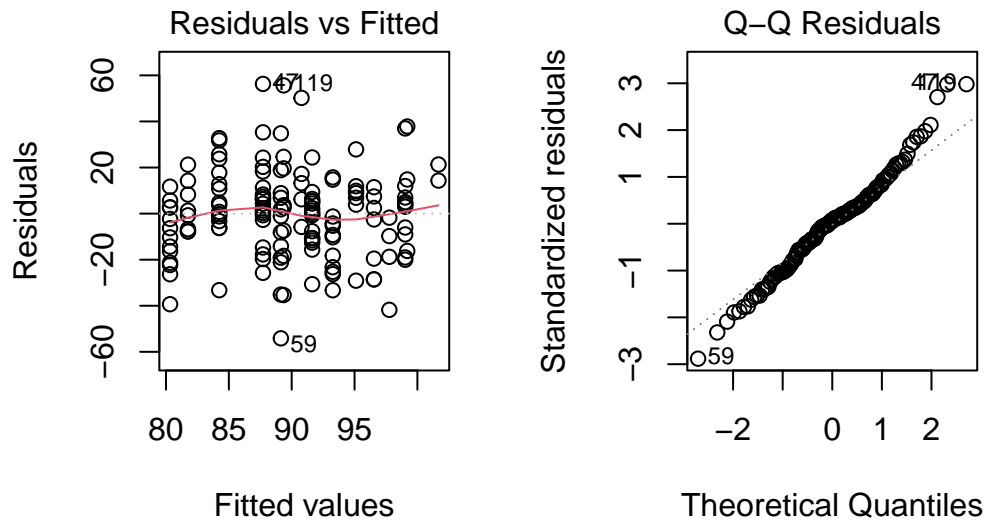
```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 14  0.8703 0.5923
      132
```

```
shapiro.test(resid(model))
```

Shapiro-Wilk normality test

```
data:  resid(model)
W = 0.98972, p-value = 0.3568
```

```
# main effects model
par(mfrow = c(1,2))
plot(main, 1:2)
```



```
# Test for age_group
leveneTest(BPAQ_score ~ age_group, data = fgame)
```

Warning in leveneTest.default(y = y, group = group, ...): group coerced to factor.

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.7209 0.1826
      144
```

```
# Test for fhrs_playing
leveneTest(BPAQ_score ~ fhrs_playing, data = fgame)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.1971 0.3148
      142
```

```
# normality
shapiro.test(resid(main))
```

Shapiro-Wilk normality test

```
data:  resid(main)
W = 0.98678, p-value = 0.1746
```

Pairwise Post-hoc Tests

```
emmeans(model, pairwise ~ age_group | fhrs_playing)$contrasts
```

fhrs_playing = 1:

contrast	estimate	SE	df	t.ratio	p.value
18 and 19 - older than 19	-4.238	5.72	132	-0.741	0.7397
18 and 19 - younger than 18	7.000	6.64	132	1.055	0.5439
older than 19 - younger than 18	11.238	6.44	132	1.744	0.1929

fhrs_playing = 2:

contrast	estimate	SE	df	t.ratio	p.value
18 and 19 - older than 19	21.529	6.65	132	3.237	0.0043
18 and 19 - younger than 18	6.875	7.71	132	0.891	0.6466
older than 19 - younger than 18	-14.654	8.00	132	-1.831	0.1635

fhrs_playing = 3:

contrast	estimate	SE	df	t.ratio	p.value
18 and 19 - older than 19	-12.250	8.13	132	-1.507	0.2910
18 and 19 - younger than 18	-24.250	9.48	132	-2.558	0.0311
older than 19 - younger than 18	-12.000	10.20	132	-1.182	0.4661

fhrs_playing = 4:

contrast	estimate	SE	df	t.ratio	p.value
18 and 19 - older than 19	-0.221	8.61	132	-0.026	0.9996
18 and 19 - younger than 18	12.351	8.61	132	1.434	0.3263
older than 19 - younger than 18	12.571	9.52	132	1.321	0.3863

fhrs_playing = 5:

contrast	estimate	SE	df	t.ratio	p.value
18 and 19 - older than 19	39.750	15.40	132	2.577	0.0296
18 and 19 - younger than 18	8.167	16.30	132	0.502	0.8703
older than 19 - younger than 18	-31.583	13.60	132	-2.322	0.0563

P value adjustment: tukey method for comparing a family of 3 estimates

Visualization of results

```
plot(emmeans(model,  
           pairwise ~ age_group | fhrs_playing)$contrasts)
```

